THE USE OF CONFIDENCE TESTING IN THE ACADEMIC INSTRUCTOR COURSE

THE SHUFORD-MASSENGILL CORPORATION
LEXINGTON
MASS. 02173

THE USE OF CONFIDENCE TESTING IN THE ACADEMIC INSTRUCTOR COURSE

WILLIE C. GARDNER, JR.

THE PROBLEM

Background

The overall mission of the Academic Instructor Course, Academic Instructor and Allied Officer School is to improve the teaching skills of selected Air Force instructors in order for them to perform more effectively in teaching assignments. AIC does not impose restrictions on the students who attend the course or require prerequisites but leaves the task of selection to the organizations who will employ the talents of the instructors after completing the course. Since the clientele that composes the normal classes is so heterogeneous, a basic fundamental core of curriculum hours is presented to all students. In order to gain information concerning the student's knowledge of the subjects to be presented and to provide guidance for faculty working with the students, a pre-course test is administered prior to student exposure to the curriculum. However, in an effort to provide maximum individualized instruction for the students, a more accurate assessment of the students' state of knowledge must be made. The students' knowledge and level of proficiency in the subject matter areas must be determined. There appear to be two basic reasons for investigating methods and techniques of measuring these "states of knowledge." These are: (1) the need to more accurately assess the students' knowledge of the curriculum upon entry into the course in order to provide maximum individualization of the curriculum, and (2) the need for more precise data upon which to base curriculum decisions.

Paper read at the 11th Annual Conference of the Military Testing Association, 15-19 September, 1969, Statler Hilton Hotel, hosted by the U.S. Coast Guard Training Center, Governors Island, New York. This research was supported in part by the Advanced Research Projects Agency of the Department of Defense, and was monitored by the Air Force Office of Scientific Research under Contract No. F44620-69-C-0068.

Student Differences

Currently, faculty members at AIC strive to meet the individual differences of students and to emphasize the need for individualized instruction with their students, who will become instructors. Most faculty advisors rely heavily upon observations of student performance, knowledge of the AIC curriculum, and experiences with the students to guide their instructional actions. Additionally, faculty advisors are provided data on each student from a reading skill test (Nelson-Denny) and a pre-course examination that is a representative sample of the AIC curriculum hours. However, the data are put to limited use since there is some question as to their accuracy and dependability. A more valid and reliable approach to pre-course testing should make it possible for the faculty to better assist the student based upon his actual needs.

More Precise Information about State of Knowledge

Although the current course pretest yields some information for the faculty advisors use in individualizing instruction, there is a need for a test that will yield additional precise information as to the students' actual levels of achievement. The current test does not employ a correction for guessing formula and, in fact, encourages the student to respond to all items on the exam, since the faculty advisors need information on all areas of the curriculum. If the test results gave a more accurate assessment of the students' actual state of knowledge (level of proficiency) the advisor could better tailor his teaching strategy in small group (seminar) activities to meet the actual needs. Although this report may not be conclusive in all aspects of its exploration, the following are a number of needs which the research at AIC can serve:

1. The need for a more accurate, concrete assessment of the students' states of knowledge. This would be in harmony with AIC's concept of meeting the students' needs through individualized instruction.

2. The need for the faculty to better meet the individual needs of students through proper guidance and appropriate teaching and reteaching strategies.

3. The need for more accurate and meaningful grades on all written tests and more accurate data for computing the final class standings.

Hypotheses

The overall problem in this study is to determine if Valid Confidence Testing techniques can provide a more accurate assessment of the students' knowledge of the subject matter and yield more useful information for the faculty than standard

choice testing. The broad hypothesis of the study is that such testing techniques can yield such information and serve as a useful tool in predicting a student's success in the course.

RESEARCH PROCEDURES

The Nature of the Data

Two separate analyses were conducted in order to include a wide range of student backgrounds and experiences and to allow comparisons of varying student groups. Class 69-A was the first to be analyzed. It consisted primarily of officers designated for either instructor or instructor-supervisor duty throughout the Air Force. The educational level of Class 69-A averaged slightly higher than the baccalaureate level. The second class to be analyzed was Class 69-B. This class consisted primarily of NCO's and a few civilians, most of whom were employed as technical instructors. Air Training Command provided the largest single group of students in that class. Education varied more widely in this class than in 69-A, with a range of 8th grade to the B.A. level. Average educational level for Class 69-B was 12.9 years.

The analysis groups consisted of 68 and 83 students, respectively. These numbers represented approximately one-half of the students in each class and were randomly selected by seminar.

Twelve variables were collected on each individual for the analysis and were provided by the records maintained by the Department of Educational Evaluation of the Academic Instructor and Allied Officer School.

Statistical Analysis

This analysis was conducted on the GE 635/645 Time-Sharing System using a remote terminal located in the Information Sciences Directorate of the Air University Institute for Professional Development. Data were input manually through the console and analyzed with both "canned" library programs and a special purpose program designed and coded by a member of the Department of Instructional Technology, AIAOS. The core of the analysis consisted of a 12 x 12 matrix of Pearson product moment correlations for each group.

Materials Used

Both classes used the materials and techniques developed by The Shuford-Massengill Corporation for taking confidence tests. These materials consisted of (1) SCoRule response aids, (2) answer sheets designed for use with the SCoRule, and (3) scoring

tables for conversion of levels of confidence into a numerical score. In addition to these materials, the students were provided with the standard multiple-choice tests administered to all AIC classes and the standard answer sheet that is used by Air University. All students were given instructions in the use of the materials via video tape and were administered a short, four-item practice test prior to the actual testing.

THE RESULTS

This section of the report will describe the major findings of the study. Because of the large number of relationships that exist in a 12 x 12 matrix many must be omitted; only the most important relationships will be discussed. If other relationships are of interest, they can be found in Appendix A.

Specific Hypotheses

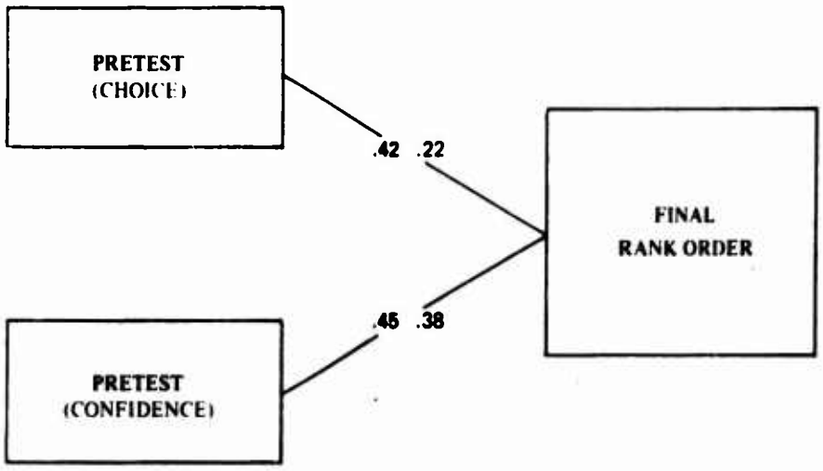
Predictive Validity

Hypothesis: Confidence tests will provide a more valid prediction of the students' success in the course and in practice teaching lessons than will choice tests.

The reasoning behind this hypothesis is that a student who can accurately evaluate his own knowledge and is confident in his knowledge of the subject matter will perform at a high level in the practice teaching lessons (PT's) which are the core of the curriculum's graded activities. The results, listed in detail in Tables 1-4, however, do not bear this out. For Class 69-A, there was no statistically significant correlation between confidence pretest scores and practice teaching efforts. For the enlisted class (69-B), the correlation was even lower. The officer class (69-A) showed a slight positive correlation between pretest confidence scores and five dependent variables: Teaching Interview PT (.15), Guided Discussion PT (.25), Lecture PT (.34), Final PT (.31), and Final Rank Order (.45). The pretest scores of the enlisted class showed the following: Demonstration Performance PT (.24)[1], Guided Discussion PT (.28), Lecture PT (.21), Final PT (.21), and Final Rank Order (.38). It should be noted that the choice pretest score for 69-A correlated with the same variables as follows: Teaching Interview (.17), Guided Discussion (.21), Lecture PT (.32), Final PT (.17), and Final Rank Order (.42). The enlisted class had the following correlations between choice pretest scores and output variables: Demonstration Performance PT (.33), Guided Discussion PT (.25), Final PT (.35), Lecture PT (.25) and Final Rank Order (.22). Except for the Guided Discussion PT, all correlations for this class between choice test scores and PT's were slightly higher than those of

[1] Demonstration Performance Method is substituted for the Teaching Interview Method with enlisted Classes. Difficulty is comparable with the Teaching Interview.

Table 1. Correlations between Pretest (Choice and Confidence) and Final Rank Order. Class 69B is in medium type.



confidence scores and PT's. We can infer then, that for officer classes the hypothesis is only partially confirmed, but not confirmed for enlisted classes. Confidence pretest scores have a slight positive correlation with practice teaching lessons and final rank order, but not significantly[2] better than choice scores for officer classes. For enlisted classes, choice scores are better predictors than confidence scores, but are at the best modest.

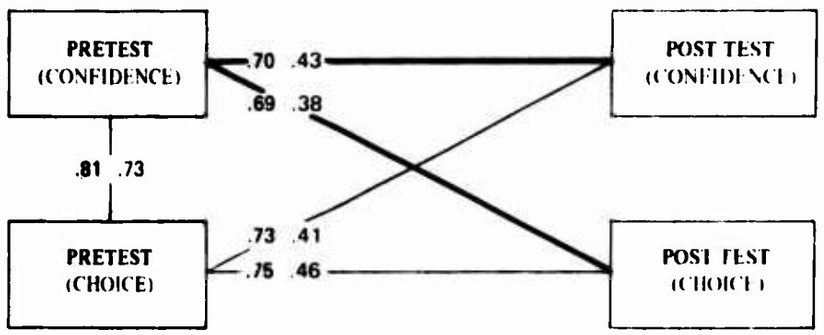
---

[2]Although normal sampling theory does not apply in this study, a test of statistical significance is applied as a general indication of the importance of the relationships. A coefficient of roughly ± .20 would be significant at the .01 level of samples of this size, indicating that there is only one chance in a hundred that the observed correlation is attributable to sampling error rather than to a real relationship of the data. This criterion is borrowed from sampling theory as a means of establishing an arbitrary cutoff point for evaluating coefficients.

In many cases the correlations may be "significant" in the statistical sense, but not very meaningful to the member or policy maker. Therefore, the reader may wish to establish his own criterion of meaningfulness. In the verbal descriptions of this report, correlations less than ± .40 are not considered very meaningful unless part of a particular suggestive pattern.

The formulas used for establishing significance levels appear in Helen M. Walker & Joseph Lev, *Statistical Inference* (New York: Holt, Rinehart & Winston, 1953) p. 251.

Table 2. Summary of Major Relationships Between Pretest Scores and Dependent Variables. Class 69B is in medium type.



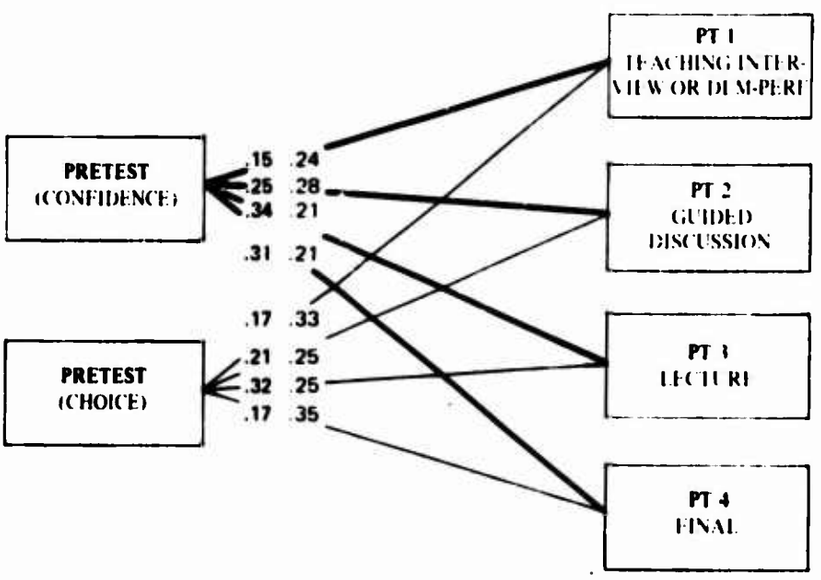


Table 3. Relationships of Reading Skill and Practice Teaching Lessons. Class 69B is in medium type.

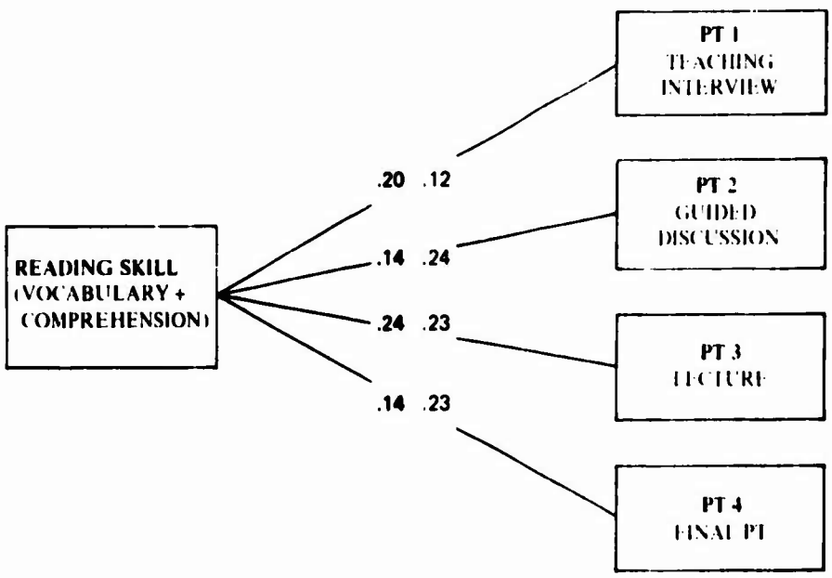
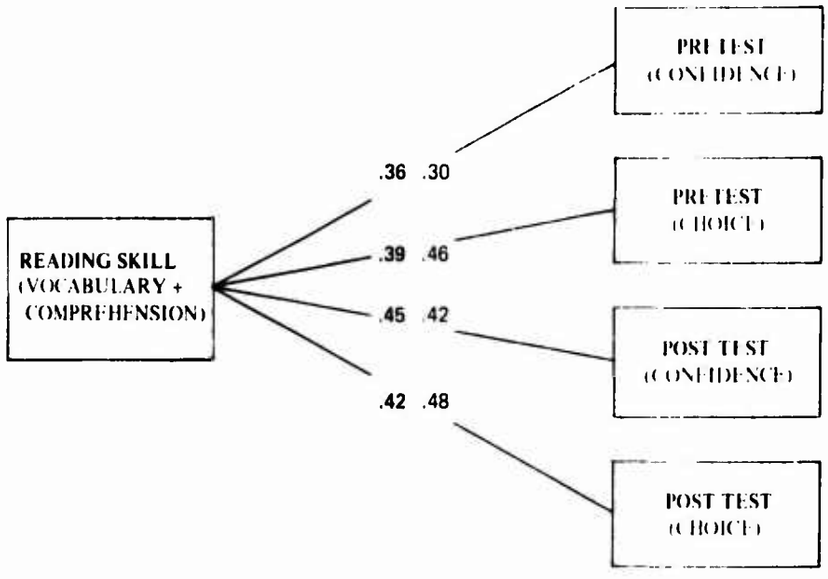


Table 4. Relationship of Reading Skill with Pre and Post Tests. Class 69B is in medium type.



Assessment of Student Knowledge

Hypothesis: The scores attained on a confidence test will be a more accurate assessment of student knowledge than will choice scores.

The basis for this hypothesis is that the scores attained by students taking a confidence test will be a more accurate reflection of their actual level of knowledge because they will more honestly and realistically analyze the items and their own knowledge before responding. The scoring system of confidence testing makes it in the best interest of the student not to guess on items of which he is not sure but rather to split his confidence among the answers he considers possible. The more confidence he has in the correct answer, the more points he receives. Also, he is not penalized if he doesn't know the answer to an item with which he is not familiar. He may split his confidence equally among all the possible answers, thus receiving some credit for his "honesty" rather than be placed in a situation in which he must guess. This also eliminates the situation in which he is penalized for guessing incorrectly when a correction for guessing formula is used. Theoretically, then, it is in his best interest not to guess on any of the items on the test. Therefore, the results should be a more accurate reflection of his actual knowledge.

The graphs listed in Tables 5 and 6 display the degree to which the students in both classes realistically evaluated information. This is based upon the percentage of times the students displayed complete confidence and were correct (Z) and the percentage of times they displayed complete confidence in an answer and were incorrect (A). When using the SCoRule, the letter "Z" represents complete confidence in an alternative while "A" indicates that there is no possibility that the alternative to which it is assigned can be the correct answer.

The more accurately the students evaluated information, the closer the lines should approach the identity line (which is an indication of ideal realism when assigning confidence to alternatives). It should be noted that with the exception of the post test for Class 69-B, each time the students were administered a test, the more nearly the results approached the identity line. With more tests, administered over a longer period of time, it is reasonable to assume that the students would become more proficient in the evaluation of their own knowledge. The post test response validity line for Class 69-B dropped 5.3% from the previous test that was administered. This may be the result of their belief that the choice test results and not the confidence scores were to be counted into their final rank order standings. Thus, they may have resorted to the normal procedure of responding to most items with complete confidence although they lacked the knowledge to justify such.

It should be noted that the officer class (69-A) consistently evaluated information more realistically. Each test that was administered achieved a higher level of response validity than did those of the enlisted class. This would tend to indicate that

Table 5. Response Validity for AIC Class 69A (All Students).

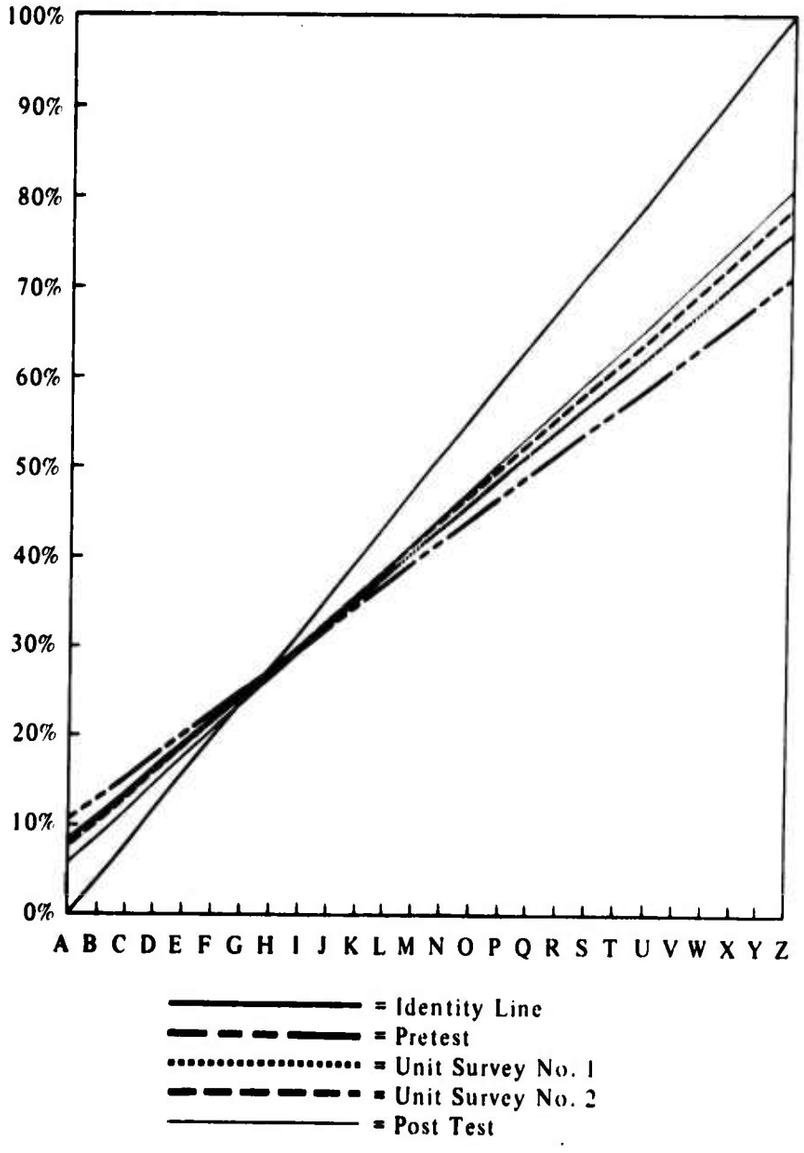
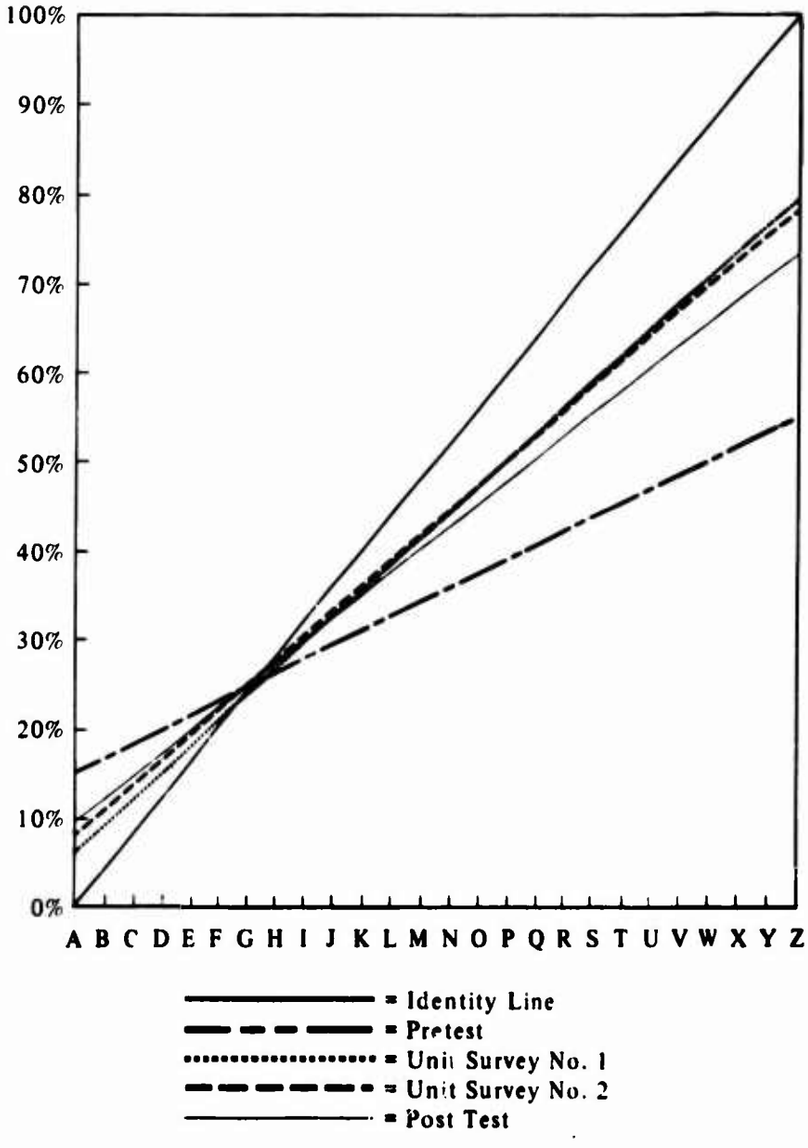


Table 6. Response Validity for AIC Class 69B (All Students).



the officer classes were better able to evaluate information than enlisted classes, perhaps because of their higher level of formal education and training.

Another measure of the accuracy of the assessment of students' knowledge is the degree to which confidence scores deviate from their choice scores. Since the students are receiving credit for partial knowledge, the scores of the confidence tests should be consistently higher than those they would receive if they were taking the test in the standard choice mode.

Tables 7 and 8 indicate the scores for the pretest and post test for Class 69-A. The confidence scores are plotted on the vertical axis, while the choice scores are on the horizontal axis. On the pretest, all but five students scored higher on the confidence test than they did when taking it as a choice test. Likewise, 59 other students achieved higher choice scores than confidence scores on the post test. These students were able to guess successfully when faced with an all-or-nothing situation. However, it should be noted that the largest deviation from confidence score was only 2.5 points and this occurred only with one student. The mean deviation of the 5-choice scores was only .675 higher than the confidence scores. This difference would appear to be insignificant, since the majority (63) of the confidence scores were much higher than those of choice.

The final index of the accuracy of measuring student knowledge reported in this study is reflected in Appendix B and C. These tables contain the comparison between choice, confidence scores and rank orders for each student on the post test for both Classes 69-A and 69-B. The officer class is characterized by many rank order inversions occurring throughout the entire distribution. As might be expected, there were generally minimal changes in position at the high and low ends of the rank order, with most significant changes occurring in the middle 90% of the cases. The most significant change in rank order occurred with student I - 1. His position changed from twenty-fifth in the group when calculated by his choice score, to eighth with his confidence score. His raw score differed from 63 (choice) to 68.90 (confidence)–5.90 points. This difference was the credit he received for partial knowledge with the confidence test, that is, he actually had more knowledge than his choice score indicated. The other extreme is illustrated by students H - 2, K - 5, B - 6, and H - 3. In each of these cases there was a significant drop in their rank order standing from that indicated by their choice scores. Students H - 2, K - 5 and B - 6 all achieved a raw score of 58 and were ranked forty-first when taking the test in the choice mode. This would indicate they all had equal knowledge and were ranked the same. However, the confidence scores were 59.00, 58.44 and 58.37 respectively, and they were ranked 51st, 52nd and 53rd, thus more accurately reflecting the differences that did exist. This differentiation resulted from the credit given by confidence testing for partial knowledge. One student, F - 2, had a raw score difference of 2.07 points, and was the only student that did not have some change in his rank order position.

Table 7. Comparison of Pretest Confidence and Choice Scores, Class 69A. Confidence raw scores are plotted on the vertical axis while the choice raw scores are on the horizontal axis.

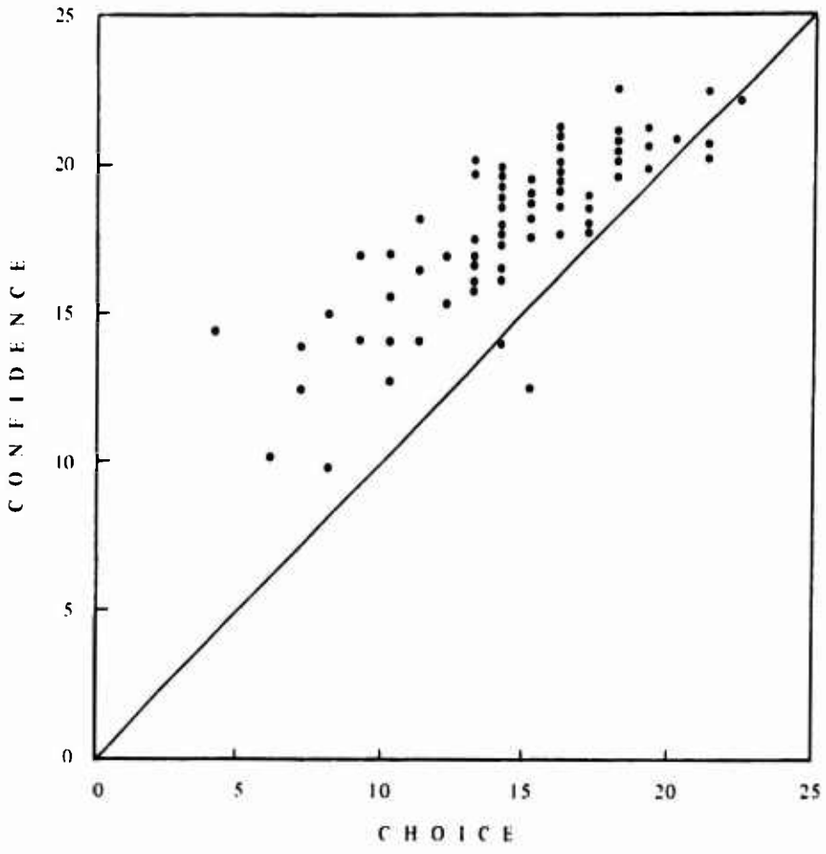
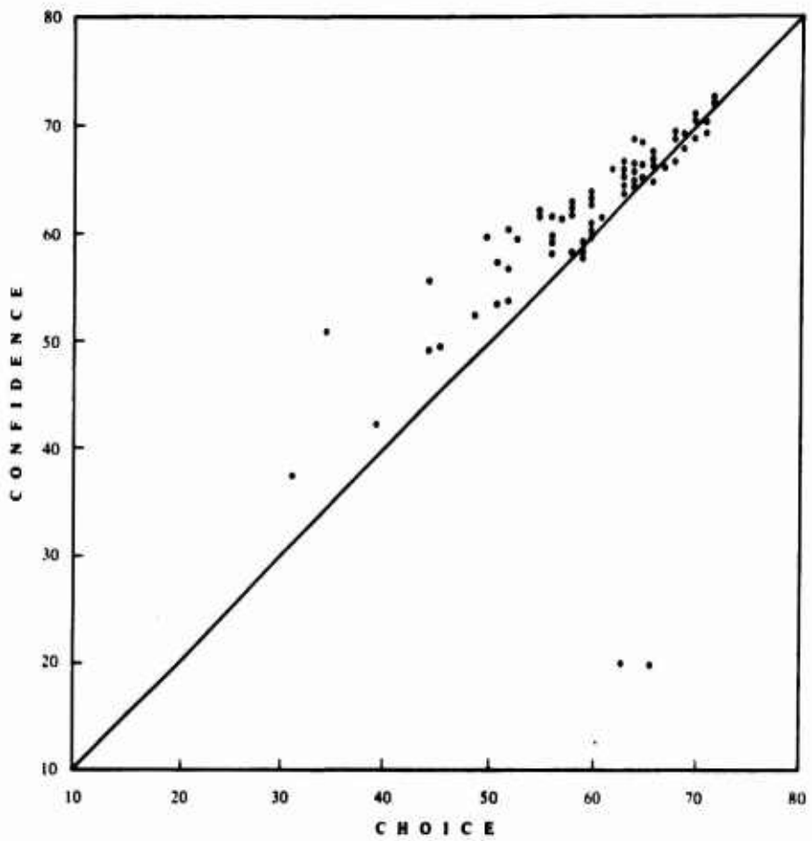


Table 8. Comparison of Post Test Confidence and Choice Scores, Class 69A. Confidence raw scores are plotted on the vertical axis while the choice raw scores are on the horizontal axis.



With the enlisted class, 69-B, the reversals in position were even more pronounced. The greatest change in rank order position occurred with student T - 5. His position changed from 60.5/84 (choice) to 22/84 (confidence), a difference of 38.5 positions. The credit for his partial knowledge gave him a raw score of 38.39 (confidence) as compared to his choice score of 29—a difference of 9.39 points. There are many other similar cases in 69-B with large reversals of position. The other extreme is illustrated with Student H - 8. He was ranked 49/84 based upon his confidence raw score of 36.20, but his choice score of 43 ranked him 4/84. The primary difference was his success at guessing the correct answer when taking the test in the choice mode. Two students, Q - 2 and N - 5, received different raw scores, but failed to change positions.

In general, the data presented in this section of the report tends to confirm the hypothesis that a more accurate assessment of the students' actual state of knowledge can be attained with the use of confidence testing. Most of the students in both classes would have been mis-classified in terms of their actual achievement had only the choice scores and accompanying rank order grades been considered, since some 98% of the students had changes in rank order positions. It seems likely that the more exposure the students have to confidence tests, the more realistically they will evaluate test items and their own knowledge, thus providing a more valid indication as to their actual level of achievement.

Reliability

Hypothesis: Confidence test results will be more reliable than choice tests since there is a significant reduction in guessing.

Perhaps the greatest single factor that affects the reliability of any test is that of guessing on the part of the students taking the test. This factor, according to J.C. Nunnally, Victor H. Noll and other authorities in evaluation, along with the clarity of the individual items, environmental conditions under which the test is administered, and the instructions that are provided the students determine, in large part, how reliable the instrument will be. Since all of the above mentioned factors can be controlled (with the exception of the guessing), for either a choice or confidence test, it is reasonable to expect that the test that least encourages the students to guess would be the more reliable instrument.

Students in Classes 69-A and 69-B were administered the pre and post tests in both choice and confidence modes. The answer sheets from the choice tests were graded with an optical scanner, and scored with a general computer scoring program. Each test was statistically analyzed by the computer to provide print-outs of the individual item analysis data (Ease and Differentiation Indices as well as alternative selections), and overall test data which included the Measures of Central Tendency, Variability and the Reliability Coefficient and computed with Kuder-Richardson

Formula 21.[3] The confidence tests were manually scored and all reliability coefficients computed using Kuder-Richardson 20. In using K-R 20 to compute the reliability coefficient, the numerical values of the levels of confidence assigned to the correct answer were used. In Table 9 the significant measures of the test are listed.

Table 9. Test Statistics for 69A and 69B Pre and Post Tests.

CLASS 69-A

	Pretest		Post Test	
	Choice	*Confidence*	*Choice*	*Confidence*
Mean	13.90	17.09	58.85	60.96
Median	14.0	17.84	57.0	62.49
Mode	13.0		58	66.00
Range	19.0	14.49	41.0	35.30
Standard Dev.	4.02	3.4	8.28	10.3
Reliability	.64	.70	.82	.90

CLASS 69-B

	Pretest		Post Test	
	Choice	*Confidence*	*Choice*	*Confidence*
Mean	11.45	14.68	32.5	35.46
Median	11	14.90	31	35.39
Mode	12	13.34	31	37.20/34.00/33.02
Range	20	11.17	25.0	23.09
Standard Dev.	3.12	3.01	5.92	7.69
Reliability	.38	.43	.81	.84

[3] For ease of computation, since it had to be performed manually, the reliabilily coefficients for all choice tests were recomputed using Kuder-Richardson Formula 20. Thus a more meaningful comparison can be made with the confidence test scores.

In all cases, there were significant increases in the reliability of the tests. This would indicate that there is some validity to the hypothesis. In addition to the data mentioned above, we can by inference ascertain that there should be a substantial increase in the reliability of the tests when they are administered using confidence techniques since there is a significant reduction of guessing. We can make this inference by observing the data presented as "States of Knowledge." The states of knowledge mentioned are derived from the level of confidence placed upon the correct answer by the student. If he has placed complete confidence on the correct answer, he is said to be "Well Informed" (W). If he places most of his confidence, but not all on the correct answer, he is classified as "Informed" (I). If he places equal confidence in either two or three of the possible answers, he is "Partially Informed" (P). When he equally splits his confidence among all the possible answers—that is, he has no ideas as to which answer might be correct—he is "Uninformed" (U). When most of a student's confidence, but not all, is placed on an incorrect answer, he is "Misinformed" (M). Finally, when he has complete confidence in an incorrect answer, he is classified as being "Completely Misinformed" (C). Thus by being able to split confidence among possible answers, we tend to eliminate guessing.

Table 10 provides the totals of the categorized states of knowledge and the percentages of students within each category for the pre and post tests for each class. From these, we can show the number of guessing "situations" which occurred for each test that were eliminated by using the confidence testing method.

With the officer class, there were 491 instances in which guessing was eliminated on the pretest. That is, the students split their confidence equally (uninformed) 274 times and divided among either two or three alternatives (partially informed) 217 times. Thus, there were 491 instances in which some degree of guessing occurred on this test when it was taken in the choice mode. Similarly, there were 438 situations in which guessing was eliminated on the post test for this class. Therefore, we can infer that the pretest for Class 69-A, when taken as a choice test was contaminated by guessing 24.2% of the time. On the post test this occurred 8.3% of the time.

The enlisted class had a slightly higher guess rate on the choice test than did the officer class. On the post test, there were 331 Uninformed responses and 291 Partially Informed responses for a total of 622 situations in which guessing occurred when taken as a choice test. This indicates that on the pretest for 69-B students guessed at 29.7% of the possible responses. The total for the post test for this class was 499, or 11.9%.

If guessing on a test can be eliminated or at least minimized as indicated above, then it is likely that the test will be a more reliable vehicle. Thus, as indicated in the initial portion of this section, and confirmed with the above data, this hypothesis appears to be valid.

Information

Hypothesis: Confidence tests yield more information that can be used in policy making for a multi-tracked curriculum than do choice tests.

A capability for diagnosing student strengths and weaknesses is needed if those assigned the responsibility of making curriculum decisions are to act intelligently to meet the individual needs of the students. With AIC attempting to individualize instruction, more than mere intuition is needed if those needs are to be met. There are only two basic sources of empirical data available at the present. These being the pretest and the Nelson-Denny Reading Test. The course pretest has been discussed earlier in this report. The reading test attempts to measure a student's reading rate, vocabulary and comprehension. As indicated in this report, the course pretest has only minimal predictive validity, and when administered as a choice or confidence test, yields little information that can be relied upon as a true indicator of how well the student will perform on practice teaching lessons or how well he will do in the course. Since the frequency of guessing is high with the choice test, too many inferences have to be made as to the actual knowledge of the subject matter. Therefore, little realistic course planning can be made using the choice pretest as a guide. The Nelson-Denny Reading Test is an excellent measure of a student's ability to read and comprehend the material, but a previous study has indicated that there is only a slight correlation between reading skills (as measured with the Nelson-Denny) and success on practice teaching lessons in the course.[4]

By virtue of the fact that confidence testing categorizes students' states of knowledge into six basic levels, it becomes possible to determine curriculum hours to be presented, and the type of teaching strategy that should be employed. Regarding curriculum hours which need to be presented, Appendix D contains the results of the pretest administered as a confidence test to Class 69-A. The table provides a breakdown of the various states of knowledge by item and the percentage of students in that classification. The left column indicates the objective that is being measured. The first seven items on the test measure hours in the Educational Fundamentals area; items 8 through 14 measure hours in the Communicative Skills area; 15 through 19 cover the Educational Methods hours; and the final 6 items measure the area of Educational Evaluation.

[4] In the study, *An Analysis of the Relationship Between Student Background Characteristics and Success at the Academic Instructor Course*, Captain Meredith W. Watts, Jr., found the correlations between reading skills and final rank order to be .29 for an officer class and .43 for an enlisted class. Additionally, the correlations between reading skill and practice teaching lessons for officer classes to be: Teaching Interview PT - .16, Guided Discussion PT - .14, Lecture PT - .27, and Final PT - .08. For enlisted classes: Demonstration Performance PT - .19, Guided Discussion PT - .23, Lecture PT - .28, And Final PT - .16.

Table 10. Summary of States of Knowledge.

	W	I	P	U	M	C
PRETEST (69A)						
Student Responses (Number)	575	358	217	274	372	301
Student Responses (Percent)	28%	17.7%	10.7%	13.5%	18.4%	14.9%
POST TEST (69A)						
Student Responses (Number)	3546	318	320	118	211	731
Student Responses (Percent)	67.6%	6%	6.1%	2.2%	4%	13.9%
PRETEST (69B)						
Student Responses (Number)	507	187	291	331	369	415
Student Responses (Percent)	24.1%	8.9%	13.9%	15.8%	17.6%	19.8%
POST TEST (69B)						
Student Responses (Number)	2371	182	322	177	245	902
Student Responses (Percent)	56.5%	4.3%	7.7%	4.2%	5.8%	21.5%

The data indicates that most students are either informed or well informed on Items 1, 2, 4 and 7. This we would expect, since most of the officers have had training and experience in Psychology (Items 1 & 2), Counseling (4), and Group Dynamics (7). It is also noted that only 15% are informed or better on the Educational Theories (Item 3), 27% in Creative Thinking (5) and 12% on Effective Thinking (6). Based upon the large percentages of students who were informed or well informed on Items 1, 2, 4 and 7, that instruction might be presented only to those students who have lower states of knowledge and offered as supplementary material to those who already are knowledgeable. It can also be noted that Items 4, 5, 6 and 7 also have a significant number of students who are either misinformed or completely misinformed. This would dictate that these students should attend instruction covering this material and that the strategy used to teach them must be one that will consider that negative transfer of learning might interfere with their receptivity. This is inferred from the large percentage of students who placed a great deal, or all, of their confidence on an incorrect alternative. If they are that positive that their answer is correct, then it may be difficult for them to receive the material without interference from previously learned material.

In the Communicative Skills area (Items 8 through 14), it is noted that more than half the group is either informed or well informed on each area measured. This indicates that it might be well to consider presenting the material on an optional basis for the informed students and as individualized remedial instruction for the remainder of the group. It appears that a sufficient number of the students are knowledgeable enough to prohibit the presentation of a "standard" package to all students.

In the Education Methods portion of the test (Items 15 through 19) the levels of knowledge are fairly well distributed among the six states of knowledge. However, there is a definite tendency for more students to be misinformed about methods and lesson planning than in the preceding area. Teaching strategy must be altered to compensate for the negative transfer mentioned concerning the material measured by Items 4, 5, 6 and 7.

In the final area, Evaluation, a different trend is evident. Most of the students have a significant amount of misinformation concerning the Characteristics of Evaluation (Item 20), and Test Construction (Items 21 and 22). As with previously mentioned areas, considerable alteration of strategy must be made. Many students are uninformed about Methods of Grading, Item Analysis, and Performance Rating areas (Items 23, 24 and 25). Nearly two-thirds (62%) indicated that they are uninformed about Grading, 75% on Item Analysis, and 27% on Performance Ratings. Once again, this information would require an alteration of the approach used by the instructor. However, this approach would be different than the one used for students who have strong misconceptions concerning the material. Experienced instructors

find it much easier to teach new material to students who have no misconceptions than to those who do. In this instance, negative transfer is not likely to hinder the students' learning.

It may be of interest to the reader to know that Item 23, when administered as a choice test item, had an Ease Index of 62%. That is, 62% of the students taking the test responded correctly. This would indicate that nearly two-thirds of the students were able to compute a T-Score prior to instruction when, in fact, only 19 students (23%) had enough knowledge of the subject to respond accordingly. This is a prime example of how misleading choice data can be when compared to how much the students actually know.

Appendix E portrays the same type data presented in the previous table. It can be noted that some of the same items that were on the pretest also appear on the post test. These are noted by the asterisks with the pretest item number in parentheses. Rather than looking at each individual item on the test, it is sufficient to say that there were significant increases in the percentages of students who were informed or higher and a marked decrease in the percentage of students who were either uninformed or misinformed on the pretest. For example, Item 24 on the pretest had a change from 4% well informed to 92% on the post test; from 75% uninformed to 4%; and from 14% misinformed to 9%. This provides the instructor or course director more information concerning the initial level of proficiency of the student and a more accurate and precise measure of the learning that has occurred as a result of instruction received in a course. Appendices F and G provide similar information for Class 69-B.

Although each individual item could be described in detail for each test, for each class, it hardly seems necessary. The data clearly indicate that significantly more information is provided by confidence tests than the standard choice tests. This material would be of great value to the individual or committee who has the responsibility for planning a multi-tracked curriculum to better meet student needs. This mode of testing also provides a greater amount and quality of information about the student than does choice testing. Thus, the instructor should be better able to adjust his teaching strategy to meet the different situations he faces.

Item Analysis Data

Hypothesis: Confidence tests will yield more data that can be used in the item analysis than will choice tests.

For any test to be consistently effective, care has to be exercised in the selection of items that will be included in that test. Periodically, subject matter will change, requiring the evaluator to update his test instrument. To do this, he must either construct new items to measure the new objectives, or rewrite the existing items that

are still usable. Selecting items to be included on a test is not a task that is accomplished intuitively, but rather is a systematic procedure that requires careful consideration be given to the item analysis data available. If the evaluator has such information as an Ease Index, Differentiation Index, and response distribution data at his disposal, the task will be greatly simplified and he is more likely to select items that will effectively measure the objectives reliably and with a high degree of validity. The above mentioned data can easily be obtained from the choice test. However, there is reason to believe that confidence tests will not only provide the same data but also additional information, perhaps with a greater degree of accuracy.

This study found that the basic formula for computing the Ease Index[5] could be used in the confidence testing situation by using the number of students who were classified as Well Informed and Informed as a substitute for the letter R in the formula. That is, these are the students who would have answered the item correctly on the choice test since they placed either all of their confidence or at least most of it on the keyed response. For the letter N, the total number of students taking the test was substituted.[6] From this data, the formula could be executed in a normal manner. It was found that the indices obtained from this method were approximately the same as those that were computed for the same items taken as a choice test, with minor deviations in both directions for the guessing involved by those students classified as Uninformed and Partially Informed.

[5] The formula used for computing the Ease Index is

$$EI = \frac{R}{N} \times 100$$

where R = the number of students responding correctly,
and N = the total number of students in the group.

[6] Consideration was given to omitting the students who were classified as Uninformed and Partially Informed from the computation, since they had not made a commitment to a single answer. However, if an accurate indication is to be made as to the Ease of the item for the entire group, the decision was made to use all students in the group when making this calculation.

The simple Differentiation Index formula[7] was treated in basically the same manner; that is, only those students that were classified as either Well Informed or Informed were included. The substitution of values is listed below. Like the computation of the Ease Index, there were only minor deviations from the Index computed for the choice test.

In spite of the finding that there were minimal differences between the indices when computed for both choice and confidence tests, it seems likely that the results obtained from the confidence computations would be a more valid assessment of the item's performance than the choice calculations. This is based on the assumption that the guessing factor was eliminated from the confidence indices. That is, the students who appeared to be guessing on the choice test (those who were Uninformed and Partially Informed) were eliminated from the calculations of the Confidence Ease and Differentiation Indices.

Additionally, the states of knowledge of the upper third of the students and the lower third could be easily plotted on a graph, thus introducing a different form of displaying the Differentiation Index for the confidence test. Table 11 portrays the graphic representation of the states of knowledge of the high and low thirds of the students for eight items on the pretest for Class 69-A. The more vertical the "W" and "I" lines, the more effective the item was in differentiating among students. Also, the longer the "W" line, the easier the item was for the two groups. (Note: It should not be interpreted as the relative ease of the item for the whole group since the middle third of the class is not considered when computing the D. I., and is omitted from the plots.) Similarly, the more horizontal the "C" and "M" lines, the more confidence placed in an incorrect answer by the lower third of the class. The shorter the "C" and "M" lines, the less misinformation possessed by these two groups. Ideally, there should be only minimal length of the "C" and "M" lines since it is highly desirable that all students successfully learn the material and evaluate their knowledge realistically. However, since all students will not be equally knowledgeable, it is desirable that the misinformation be identified. Usually, misinformation is more likely to occur with those students who achieve the lower scores and, in turn, have mastered less subject matter than those with higher scores.

[7] The formula used to compute the Differentiation Index was:

$$DI = \frac{H - L}{N/3}$$

where H = the number of students in the high one-third of the group who were classified as W or I, and L = those in the low one-third who were classed as W or I, N = the total number of students in the group, and 3 = Number of Criterion Groups.

Table 11(a). States of Knowledge of the Top and Low Criterion Groups.

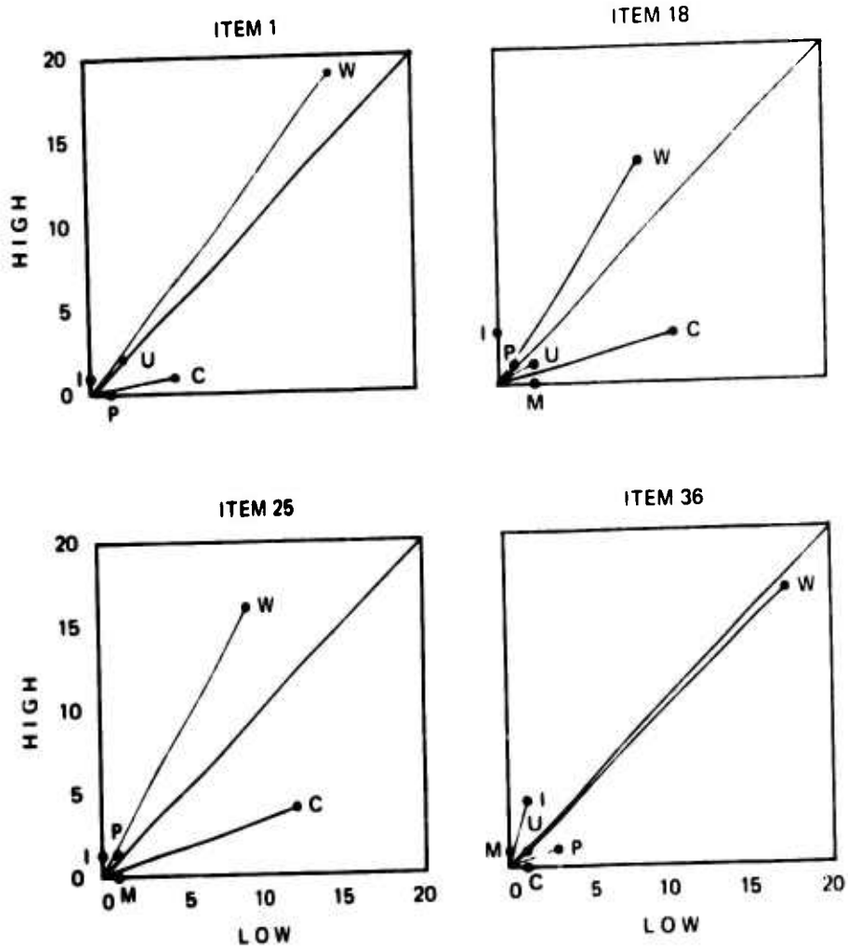
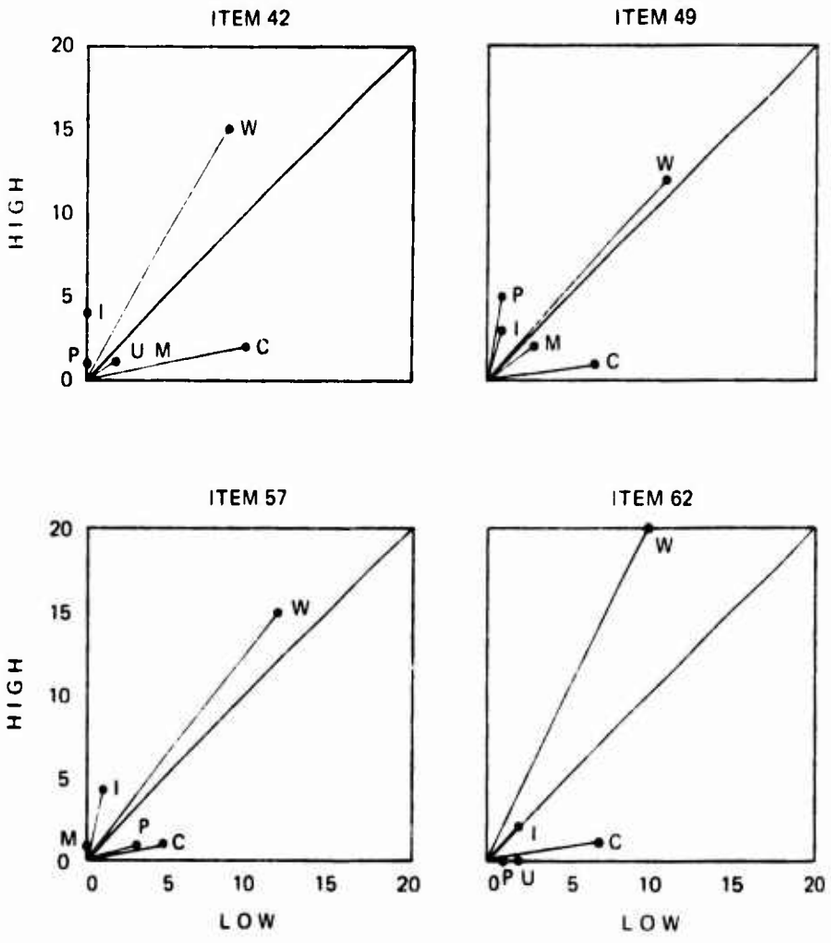


Table 11(b). States of Knowledge of the Top and Low Criterion Groups.



Thus, with the ability to compute two basic measures of an item's effectiveness, the Ease and Differentiation Indices, along with the ability to display the Differentiation in states of knowledge in a graphic manner, it appears that the hypothesis is partially confirmed. However, an evaluator also uses the analysis of response effectiveness in order to determine how effectively the distracting alternatives are performing. With the standard choice test, this is done by plotting the number of responses for each item by alternative and by criterion group. This tally for a typical item appears as follows:

Item 25 Ease Index: 59%, Differentiation Index: .48

Alternatives	A	B*	C	D
High 1/3	1	20	2	2
Middle 1/3	3	16	3	3
Low 1/3	7	8	5	5
Totals:	11	44	10	10

With this data, he can compute the Ease and Differentiation indices by using the numbers plotted for the high and low criterion groups in the alternative "B" column along with the total number of students in the group. Additionally, he can observe the quality of the distractors by visual inspection. When he finds that no one has selected a particular alternative, he can assume that it is not plausible and replace it with one which is.

However, from the above data the evaluator is unable to determine if a student even considered selecting an alternative other than his actual choice. Confidence testing can provide this information, thus indicating that the alternative was actually more plausible than it appeared on the choice test.

Since a student can assign some degree of confidence to all possible answers he feels may be correct, an analysis of these levels of confidence can be made. This analysis will provide a better indication of how plausible the distractor actually was. Tables 12 through 17 contain data that indicates how many times each level of confidence was used and the cumulative frequency for the various levels. Additionally, the cumulative frequency has been plotted graphically for each of the items, providing a visualization of the relative effectiveness of each of the alternatives. For example, Item 1 indicates that most of the 68 students in 69-A had a great deal of confidence in the correct answer (alternative number 1). With alternative 2, most students (57) assigned no confidence to it as being possibly correct, while three were positive it was the correct answer. Were this item considered from a choice test standpoint, only the three students who assigned complete co ıfidence to this alternative would have appeared on the tally mentioned previously. There is a possibility that one, two or all three of the students who assigned half their confidence (M level) would have selected this as their answer, depending upon how they guessed. However, none of the remaining five students would have given any indication that they had even considered this as being possibly correct, since they had more confidence in another alternative. Thus, we find that the distractor is working more effectively than one might imagine.

Similarly, alternative 3 had a slightly better distracting ability than did alternative 2, since two fewer students were able to eliminate this as a possibility. There were three students who placed some degree of confidence on the alternative that would not have been recognized with choice testing. Likewise, four responses for alternative 4 would not have been discovered.

In a more dramatic representation, Item 57, (Tables 15-17), clearly indicates how the test constructor might be misled when determining the effectiveness of an item's distractors. With the keyed (correct) response (1) there were nine students who had some confidence in the response, but would not have selected it by choice. This would indicate they at least had enough knowledge concerning the subject matter to consider it as being a possibility. Also, nine students divided their confidence equally between the keyed response and another alternative. Depending upon how successful they were in guessing, the Ease Index for this item could vary by 23%, thus giving an inaccurate indica ion of its effectiveness and relative ease. In alternative 3, 13 students indicated degrees of confidence that would not have been identified. Alternative 4 had 11 that would not have been identified.

Table 12. Number of Times Each Degree of Confidence Used for Item 1 (post test).

Level of Confidence	Alternative Number 1	Alternative Number 2	Alternative Number 3	Alternative Number 4
A (None)	17	57	55	56
B	1	1		1
C		1		1
D		2	1	1
E			1	
F				
G				
H		1		
I	1		1	1
J				
K				
L				
M	3	3	2	2
N				
O				
P				1
Q				
R	2			
S				
T				
U			1	
V	1			
W			1	
X				
Y				
Z (Complete)	43	3	6	5
TOTAL	68	68	68	68

Table 13. Frequency Distribution of Item 1.

Level of Confidence	Alternative Number 1	Alternative Number 2	Alternative Number 3	Alternative Number 4
Z (Complete)	43	3	6	5
Y				
X				
W			7	
V	44			
U			8	
T				
S				
R	46			
Q				
P				6
O				
N				8
M	49	6	10	
L				
K				
J				
I	50		11	9
H			7	
G				
F				
E			12	
D		9	13	10
C		10		11
B	51	11		12
A (None)	68	68	68	68

Table 14(a). Total Number of Students (Class 69A) Who Expressed at Least Some Confidence in Each Response to Item 1 (post test).

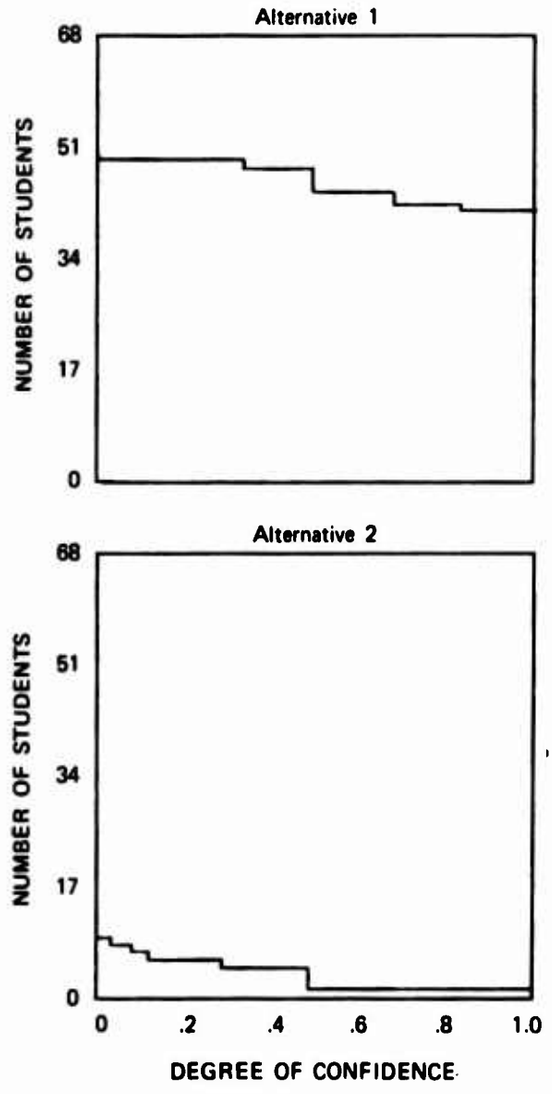


Table 14(b). Total Number of Students (Class 69A) Who Expressed at Least Some Confidence in Each Response to Item 1 (post test).

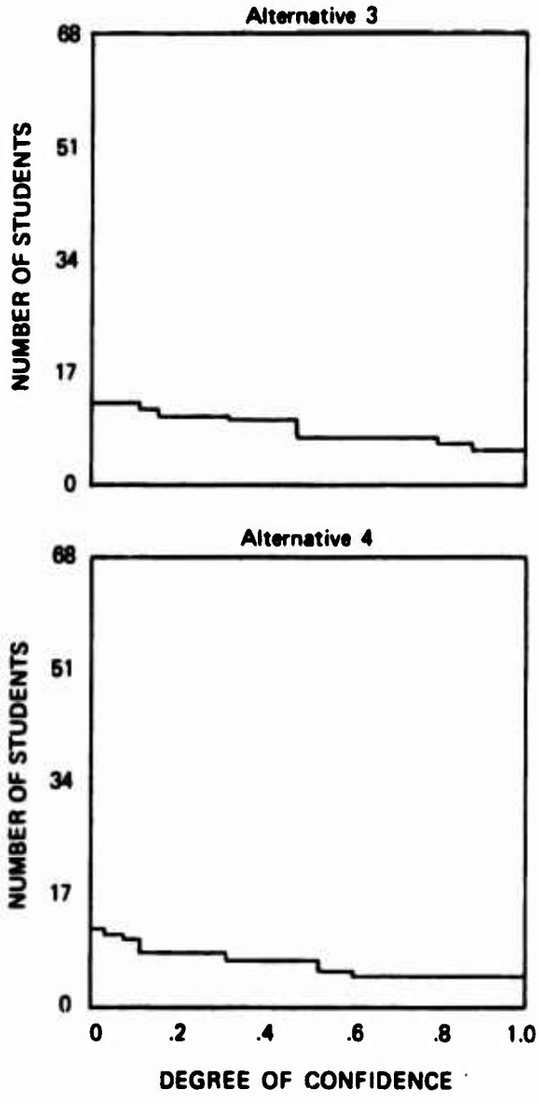


Table 15. Number of Times Each Degree of Confidence Used for Item 57 (post test).

Level of Confidence	Alternative Number 1	Alternative Number 2	Alternative Number 3	Alternative Number 4
A	12	54	46	51
B		1	2	1
C		1	2	2
D	1			
E	1		2	3
F				
G	3	3	3	4
H				
I	2	1	2	1
J	1		1	
K	1		1	
L				
M	9	2	4	3
N				
O				
P				
Q	1			
R	1			
S				
T	2			
U			1	
V				1
W	1			
X				
Y	1			
Z	32	6	4	2
TOTAL	68	68	68	68

Table 16. Cumulative Frequency Distribution for Item 57.

Level of Confidence	Alternative Number 1	Alternative Number 2	Alternative Number 3	Alternative Number 4
Z	32	6	4	2
Y	33			
X				
W	34			
V				3
U			5	
T	36			
S				
R	37			
Q	38			
P				
O				
N				
M	47	8	9	6
L				
K	48		10	
J	49		11	
I	51	9	13	7
H				
G	54	12	16	11
F				
E	55		18	14
D	56			
C		13	20	16
B		14	22	17
A	68	68	68	68

Table 17(a). Total Number of Students (Class 69A) Who Expressed at Least Some Confidence in Each Response to Item 57 (post test).

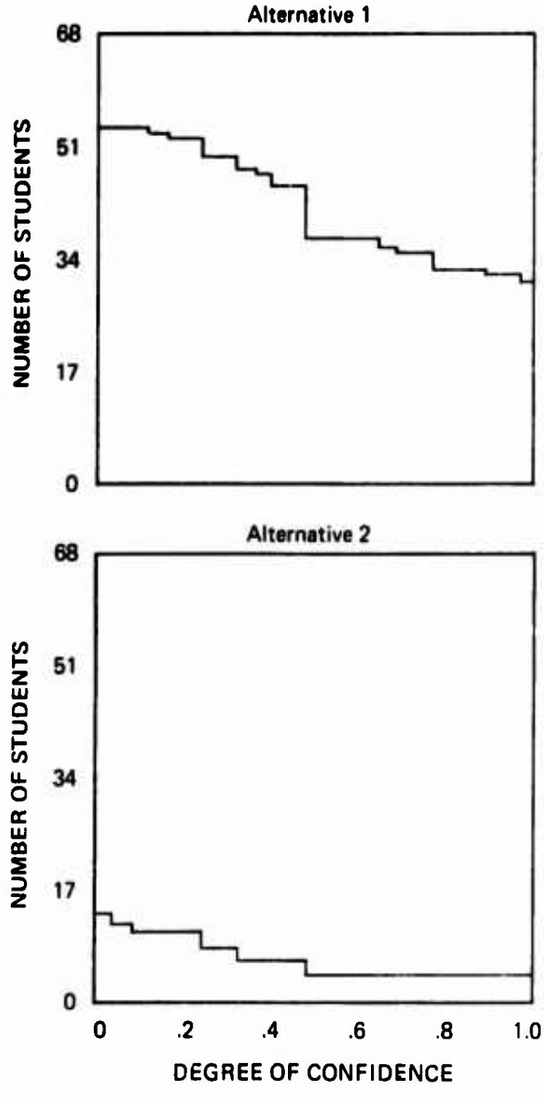
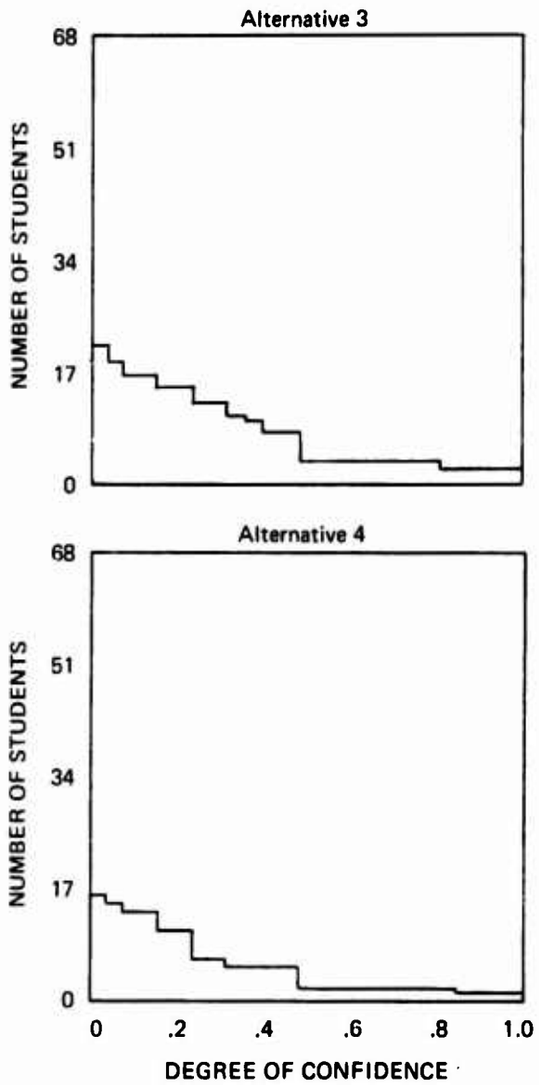


Table 17(b). Total Number of Students (Class 69A) Who Expressed at Least Some Confidence in Each Response to Item 57 (post test).



Still another visual representation can be made of the data provided by a confidence test. This representation involves the same basic technique mentioned above, that is, using the cumulative frequencies of the levels of assigned confidence, except that the high third of the group and the low third have been combined on each graph. This provides another means of observing the differences in the responses of the better students and the weaker ones. Computing the Differentiation Index, as mentioned earlier, provides a numerical indication of how effectively the items differentiate among students in designed criterion groups. The charts contained in Tables 18 and 19 provide a visualization of that information and also display the differences in levels of confidence assigned to each of the distractors by each of the groups. For example, Item 1 on Table 18 shows that there were substantial differences in the levels of confidence assigned to the correct answer (alternative 1). Alternative 2 indicates that more of the weaker students placed more confidence in this answer than did the better students. For the lower one-third, it sounded more plausible than it was for the upper third. Alternative 3 discloses that approximately the same numbers of top and bottom students placed about the same degrees of confidence in the answer. The last alternative for this item clearly shows that the lower third found this choice more plausible than did students in the top third. The reader will note that no student in the top third place more than 50% of his confidence in that response, while four students in the lower third placed complete confidence in it as being the correct answer. With this manner of representing the differences in assigned confidence levels, the greater the area between the upper levels of the lines, the greater the differentiation between groups. With this form of data reporting, it is possible to compute some form of "differentiation index" among the alternatives.

Ideally, there should be a significantly large area between the assigned levels of confidence for the top and bottom groups on the correct answer, with the top group greater in number than that of the lower. With the distractors, the lower group's level of confidence line should exceed the top group's, since the lower group has likely failed to master the subject matter and would most likely display greater confidence in an incorrect alternative than would the students who were placed in the top group based upon their overall test score. Item 25 on Table 18 illustrates this point. Alternative 3 appears to be an excellent distractor, since it caused a sufficient number of lower students to place complete confidence in it. The keyed response, alternative 4, clearly indicates that most of the students in the top third assigned a high degree of confidence to this answer, while approximately one-half of the lower third assigned corresponding levels of confidence to it.

Each item can be covered in detail by the reader, but it is evident that this method of analyzing the effectiveness of the alternatives provides more information than can be obtained with a normal choice test. The above information, along with the ability to compute the Ease and Differentiation Indices with more validity and to display the findings in other than numerical terms, tends to confirm the hypothesis of this part of the study.

Table 18(a). Comparison of Confidence Distributions for the Four Alternatives of Item 1 and Item 18.

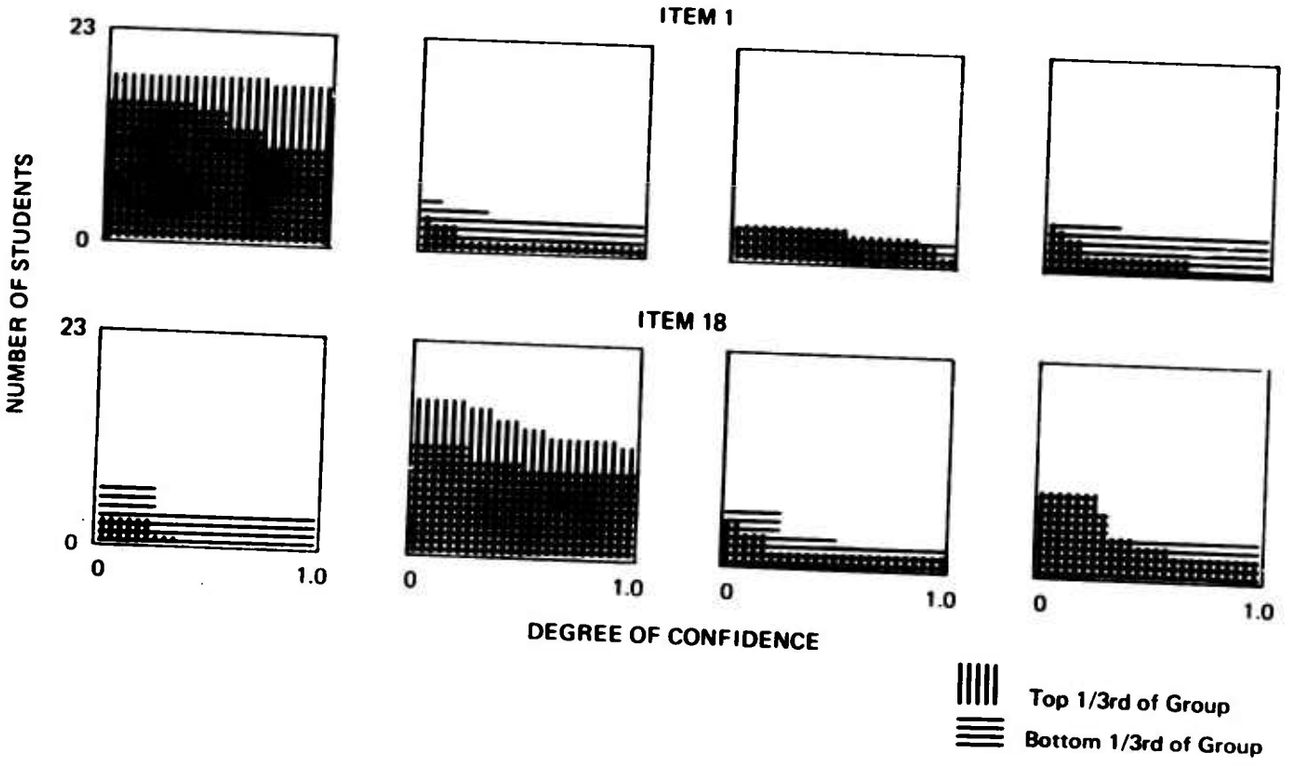


Table 18(b). Comparison of Confidence Distributions for the Four Alternatives of Item 25 and Item 36.

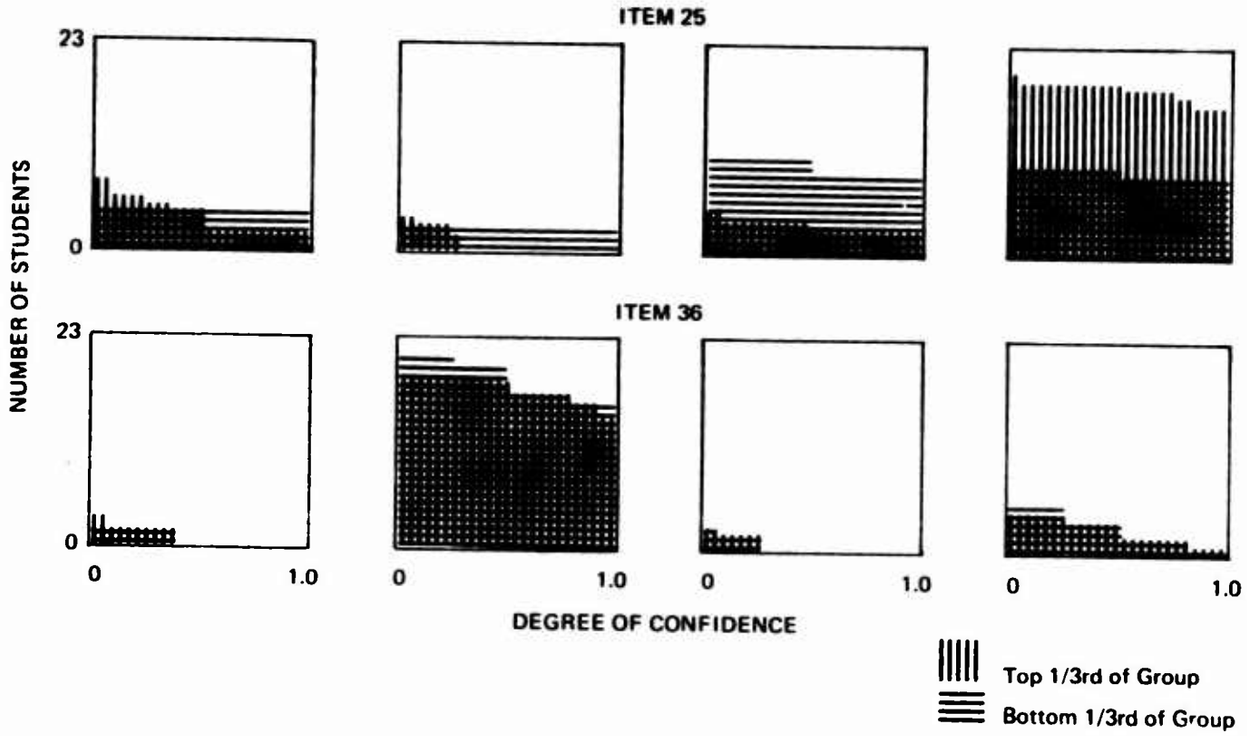


Table 19(a). Comparison of Confidence Distributions for the Four Alternatives of Item 42 and Item 49.

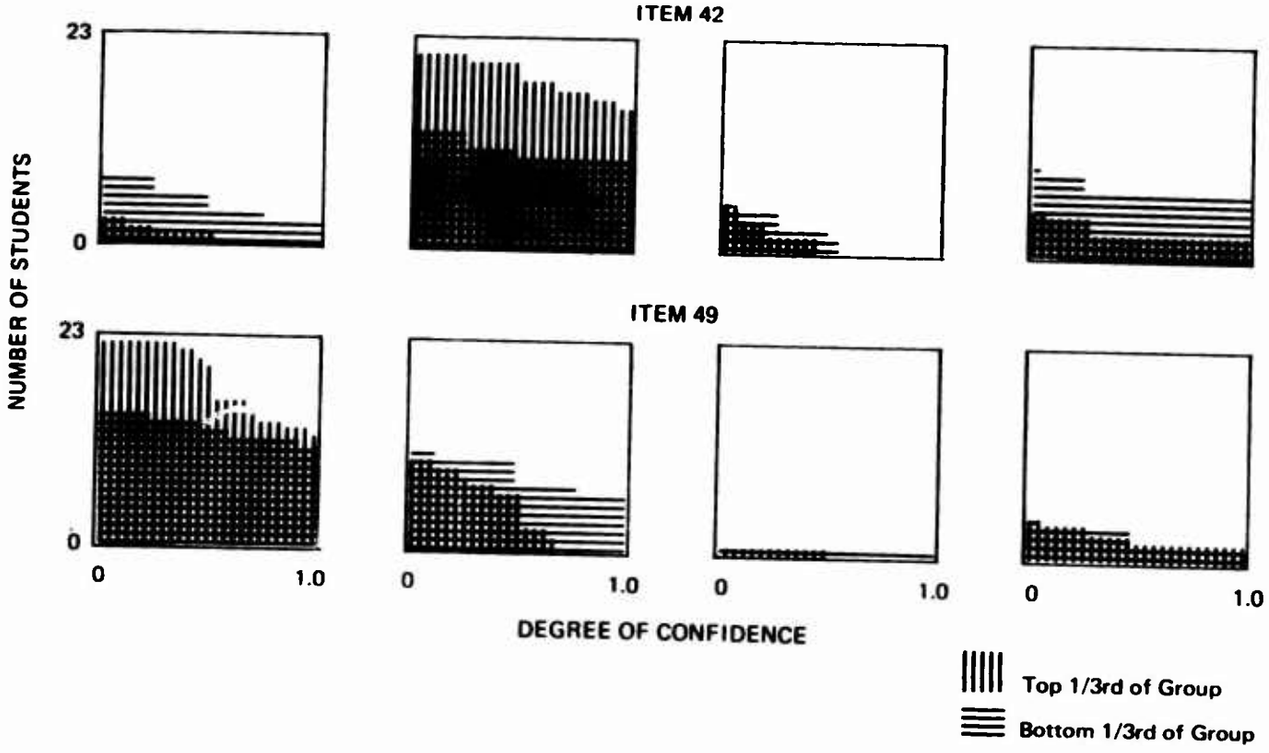
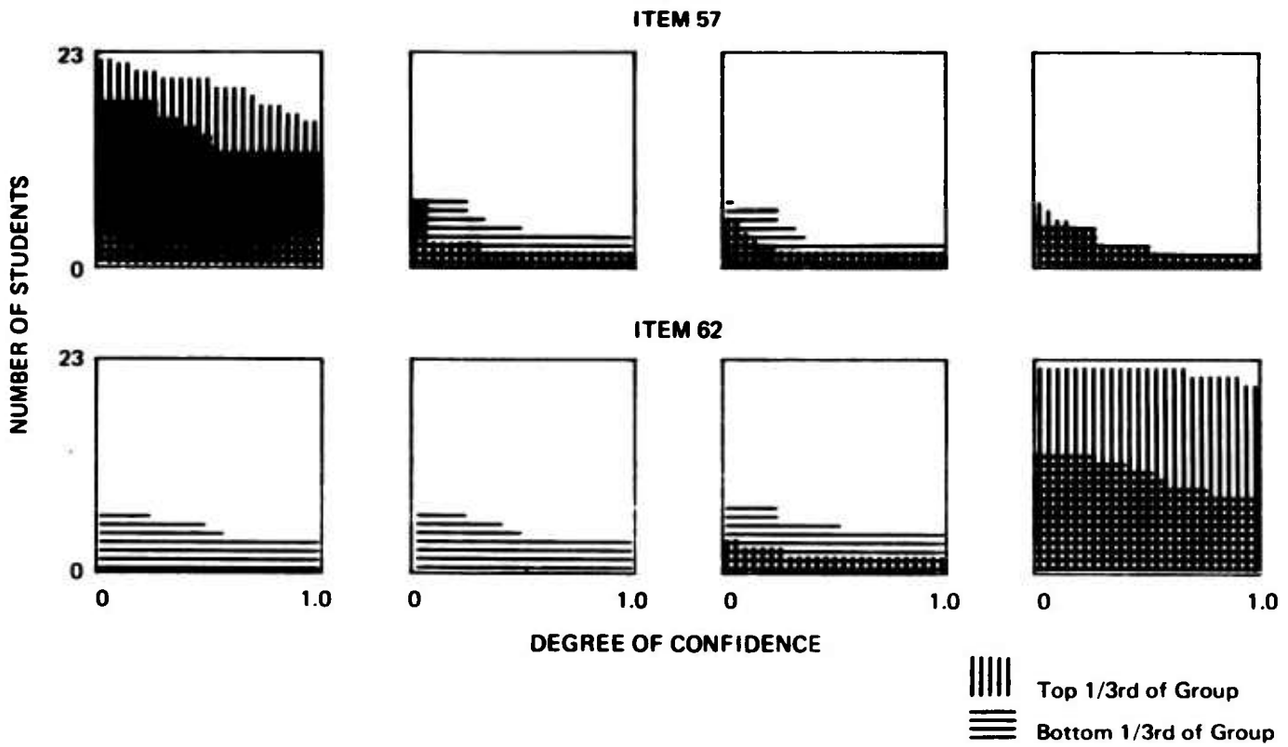


Table 19(b). Comparison of Confidence Distributions for the Four Alternatives of Item 57 and Item 62.



CONCLUSION

Summary of Major Relationships

The problems stated at the outset were: to discover whether information provided by confidence testing would result in a more accurate assessment of student knowledge so more personal individualized instruction could be presented, and whether sufficient precise data could be attained to be of significant aid in curriculum planning. Analysis has shown that much qualitative and quantitative information is provided when employing the methods and materials associated with confidence testing. A compendium of those findings follows:

Predictive Validity appears to be positive, but not significantly high enough to forecast a student's performance on practice teaching lessons or his final standing in the class. This is more strongly associated with the officer class than with the enlisted class. Choice pretest scores have a slightly higher predictive value than do confidence scores for the enlisted class. Neither choice nor confidence pretest scores correlate significantly enough with success at AIC to warrant use as predictive devices.

The *Assessment of Student Knowledge* tends to be more accurate when using confidence tests. Data presented in this section of the report clearly indicate that most students would have been misclassified as to their actual levels of knowledge if only choice test results had been used. The frequency of reversals of rank order position indicate that when a student's actual state of knowledge is more accurately assessed his position within the group changes, thus reflecting more accurate differentiation. There are strong reasons to believe that students tend to evaluate their knowledge more carefully and critically when using confidence tests, and the more often they are tested with the materials, the more realistic this evaluation becomes.

Reliability of test instruments seems to be improved when using confidence tests. In the one instance when the actual reliability coefficient was computed, a significant increase was noted. This, coupled with the fact that significant numbers of guessing situations were eliminated by using the confidence test, indicates that the accuracy and reliability of the instrument is likely to be increased.

Confidence tests provide more *Information* about the students than do choice tests. The ability of the confidence test to categorize states of knowledge for each item on a pretest would make multi-tracking decisions easier for curriculum planners. The ability of the confidence test to identify the various levels of proficiency for each student in each area of the curriculum provides an invaluable aid to the instructor in planning his teaching strategy. By being able to identify students who need

supplementary material, who need reinforcement, who are apt to have difficulties because of negative transfer of learning and who know nothing about the material to be presented, the instructor can more readily adapt to meet the needs of each individual.

Better *Item Analysis Data* is provided with confidence tests than with the standard choice tests. Both methods of testing provide the data to compute the basic measures of an item's effectiveness, but confidence tests provide a more accurate index since responses involving guessing are not a factor in the computations. The ability of the confidence test to identify those students who cannot rule out the incorrect responses to an item provides valuable data to the individual constructing or analyzing the test. A more accurate assessment can be made of the distractor's effectiveness.

Suggestions for Further Research

The analysis reported here is basic and exploratory in nature and, although it does test several specific hypotheses, it needs more statistical refinement. Modifications in design, statistical analysis, and inclusion of new variables should be considered for further research relating to the use of confidence testing in military situations. The following areas are suggested for future research:

1. The use of the confidence testing materials and techniques with the Allied Officers who attend the Allied Officer Familiarization Course (AOFC). This area appears to be one in which these techniques could be especially useful with those students who are having some difficulty with the written portion of the English language.

2. Continued research with different variables to determine if the technique can be used to measure all aspects of affective as well as cognitive learning areas. The ability of a written test instrument to measure such things as attitude, interests, and appreciations has not been successfully developed, thus opening the possibility for additional research.

3. Continued research with confidence testing to determine the amount of repetition that is needed if the test's validity and reliability is to be maintained. Theoretically, if all items were sound qualitatively repetition would be eliminated without decreasing the reliability of the instrument.

4. Research oriented toward the use of confidence tests in performance situations. The elements of exercises such as practice teaching lessons could be tested prior to the student performance to determine if the written test can predict success in actual performance.

REFERENCES

FROELICH, F.J. A Simple Index of Test Reliability, *Journal of Educational Psychology*, 1941, **32**:381-385.

NOLL, V.H. *Introduction to Educational Measurement*, Boston:Houghton Mifflin Company, 1965.

NUNNALLY, J.C. *Educational Measurement and Evaluation*, New York:McGraw-Hill, 1964, **68**:403-408.

SHUFORD, E.H., Jr., ALBERT, Arthur, and MASSENGILL, H.E. Admissible Probability Measurement Procedures, *Psychometrika*, 31:No. 2 (June, 1966), pp. 125-145.

SHUFORD, E.H., Jr., and MASSENGILL, H.E. The Effect of Guessing on the Quality of Personnel and Counseling Decisions. Report 5 of The Shuford-Massengill Corporation, 1966.

SHUFORD, E.H., Jr., and MASSENGILL, H.E. A Report on the Degree of Confidence in Student Testing. Report 11 of The Shuford-Massengill Corporation, 1968.

WATTS, M.W., Jr. An Analysis of the Relationship Between Student Background Characteristics and Success at the Academic Instructor Course, Unpublished research report, Maxwell AFB, Alabama.

Appendix A: Summary of the Relationships between Variables

	1	2	3	4	5	6	7	8	9	10	11	12
1. Pretest (Conf.)	**1.00**	**.29**	**.81**	**.36**	**.15**	**.25**	**.34**	**.32**	**.31**	**.45**	**.69**	**.70**
	1.00	.36	.73	.30	.24	.28	.21	.31	.21	.38	.38	.43
2. Formal Education		**1.00**	**.17**	**.20**	**.32**	**.02**	**.38**	**.33**	**.20**	**.39**	**.29**	**.30**
		1.00	.42	.30	.10	.08	.001	.07	.04	.001	.22	.19
3. Pretest (Choice)			**1.00**	**.39**	**.17**	**.21**	**.32**	**.30**	**.17**	**.42**	**.75**	**.73**
			1.00	.46	.33	.25	.25	.25	.35	.22	.46	.41
4. Reading Skill				**1.00**	**.20**	**.14**	**.24**	**.24**	**.14**	**.25**	**.42**	**.45**
				1.00	.12	.24	.23	.25	.23	.39	.48	.42
5. PT 1 (Teach Int. or Dem-Perf.)					**1.00**	**.36**	**.31**	**.74**	**.46**	**.67**	**.13**	**.18**
					1.00	.45	.42	.79	.50	.70	.14	.0
6. Guided Discussion PT						**1.00**	**.36**	**.75**	**.34**	**.61**	**.24**	**.28**
						1.00	.39	.74	.63	.72	.46	.41
7. Lecture PT							**1.00**	**.75**	**.42**	**.67**	**.37**	**.39**
							1.00	.81	.52	.69	.21	.18
8. Average of PTs 1, 2, 3								**1.00**	**.53**	**.86**	**.33**	**.38**
								1.00	.70	.89	.33	.27
9. Final PT									**1.00**	**.70**	**.15**	**.22**
									1.00	.80	.31	.20
10. Final Rank Order										**1.00**	**.50**	**.52**
										1.00	.51	.47
11. Post Test (Choice)											**1.00**	**.94**
											1.00	.81
12. Post Test (Confidence)												**1.00**
												1.00

Appendix B: Post Test Results (Class 69A)

Student Number	Post Test Score		Rank Order	
	Confidence	*Choice*	*Confidence*	*Choice*
I - 2	72.30	71	1	1.5
O - 2	71.71	71	2	1.5
K - 4	70.69	69	3	6.0
G - 5	70.44	69	4	6.0
N - 4	70.00	70	5	3.5
H - 7	69.03	67	6	11.0
N - 3	69.00	70	7	3.5
I - 1	68.90	63	8	25.0
G - 4	68.89	69	9	6.0
L - 3	68.81	67	10	11.0
J - 6	68.66	68	11	8.5
J - 7	68.06	64	12	20.5
G - 3	67.87	65	13	16.5
J - 2	67.46	68	14	8.5
F - 1	66.89	67	15	11.0
I - 5	66.83	62	16	30.0
L - 4	66.77	65	17	16.5
I - 4	66.74	63	18	25.0
N - 2	66.25	64	19	20.5
K - 2	66.00	66	20.5	13.5
N - 5	66.00	66	20.5	13.5
O - 7	65.89	61	22	33.0
G - 6	65.88	65	23	16.5
J - 3	65.70	63	24	25.0
G - 1	65.68	62	25	30.0
L - 6	64.84	65	26	16.5
L - 5	64.68	64	27	20.5
N - 7	64.52	62	28	30.0
O - 5	64.42	64	29	20.5
F - 2	64.07	62	30	30.0
B - 4	64.00	63	31	25.0
B - 3	63.86	63	32	25.0
O - 6	63.56	59	33	37.0
L - 2	62.49	59	34	37.0

Appendix B: Post Test Results (Class 69A)

Student Number	Post Test Score		Rank Order	
	Confidence	*Choice*	*Confidence*	*Choice*
O - 1	62.31	57	35	44.0
I - 7	62.23	57	36	44.0
J - 4	62.07	62	37	30.0
H - 5	62.00	54	38	51.5
K - 6	61.94	54	39	51.5
H - 4	61.84	60	40	34.0
L - 7	61.37	56	41	46.0
F - 5	60.89	55	42	48.5
H - 1	60.73	59	43	37.0
J - 1	60.54	59	44	37.0
J - 5	60.30	51	45	53.0
G - 7	60.00	59	46	37.0
I - 6	59.94	49	47	59.0
H - 6	59.90	55	48	48.5
O - 3	59.42	55	49	48.5
I - 3	59.05	52	50	53.0
H - 2	59.00	58	51	41.0
K - 5	58.44	58	52	41.0
B - 6	58.37	58	53	41.0
H - 3	58.10	57	54	44.0
B - 7	58.08	55	55	48.5
K - 2	57.26	50	56	57.5
B - 1	56.93	51	57	55.0
F - 6	55.36	44	58	63.5
B - 2	53.95	51	59	55.0
G - 2	53.69	50	60	57.5
L - 1	52.59	48	61	60.5
K - 3	50.73	34	62	67.0
N - 6	49.98	45	63	62.0
G - 4	49.36	44	64	63.5
N - 1	48.52	48	65	60.5
F - 3	42.03	39	66	65.0
O - 4	37.42	31	67	68.0
B - 5	37.00	37	68	66.0

Appendix C: Post Test Results (Class 69B)

Student Number	Post Test Score		Rank Order	
	Confidence	*Choice*	*Confidence*	*Choice*
H - 5	45.84	44	1	1.5
E - 6	45.37	43	2	4.0
E - 8	44.43	44	3	1.5
H - 4	44.16	41	4	10.0
I - 5	44.03	43	5	4.0
J - 2	43.87	40	6	14.0
I - 3	43.21	40	7	14.0
P - 5	42.84	42	8	7.0
I - 2	42.73	39	9	18.5
Q - 5	42.00	42	10	7.0
E - 3	41.96	39	11	18.5
L - 2	41.84	41	12	10.0
Q - 7	41.69	42	13	7.0
L - 8	41.64	41	14	10.0
I - 6	40.87	35	15	30.0
Q - 6	40.39	40	16	14.0
H - 6	40.20	40	17	14.0
G - 3	40.00	40	18	14.0
M - 2	39.29	39	19	18.5
M - 1	39.00	39	20	18.5
E - 2	38.58	32	21	48.5
T - 5	38.39	29	22	60.5
H - 2	38.33	29	23	60.5
G - 6	38.27	35	24	30.0
E - 5	38.20	37	25	24.5
N - 6	38.18	36	26	26.5
I - 8	38.02	33	27	41.5
M - 3	38.00	38	28	22.0

Appendix C: Post Test Results (Class 69B)

Student Number	Post Test Score		Rank Order	
	Confidence	*Choice*	*Confidence*	*Choice*
J - 3	**37.58**	**38**	29	22.0
E - 1	**37.42**	**36**	30	26.5
J - 4	**37.35**	**33**	31	41.5
Q - 1	**37.20**	**38**	32.5	22.0
Q - 3	**37.20**	**37**	32.5	24.5
G - 4	**36.79**	**34**	34	35.0
Q - 2	**36.74**	**34**	35	35.0
L - 1	**36.62**	**35**	36	30.0
E - 7	**36.58**	**33**	37	24.5
M - 5	**36.52**	**32**	38	48.5
T - 2	**36.41**	**34**	39	35.0
H - 8	**36.20**	**43**	40	4.0
G - 8	**35.79**	**30**	41	56.0
J - 1	**35.39**	**34**	42	35.0
T - 6	**35.37**	**30**	43	56.0
N - 1	**35.18**	**25**	44	73.5
L - 3	**35.14**	**32**	45	48.5
H - 1	**34.67**	**32**	46	48.5
P - 1	**34.52**	**33**	47	41.5
M - 6	**34.24**	**32**	48	48.5
H - 7	**34.00**	**33**	49.5	41.5
P - 4	**34.00**	**34**	49.5	35.0
T - 3	**33.90**	**31**	51	52.5
N - 4	**33.84**	**35**	52	30.0
T - 1	**33.76**	**17**	53	83.0
M - 4	**33.57**	**35**	54	30.0
L - 6	**33.22**	**33**	55	41.5
J - 8	**33.15**	**25**	56	73.5

Appendix C: Post Test Results (Class 69B)

Student Number	Post Test Score		Rank Order	
	Confidence	*Choice*	*Confidence*	*Choice*
J - 5	33.02	31	57.5	52.5
J - 6	33.02	24	57.5	78.0
T - 7	32.77	27	59	67.5
J - 7	32.74	27	60	67.5
P - 2	32.68	33	61	41.5
I - 7	32.66	16	62	84.0
P - 6	32.59	30	63	56.0
N - 5	32.49	28	64	64.0
Q - 4	32.00	32	65	48.5
M - 7	31.84	30	66	56.0
I - 1	31.31	20	67	82.0
I - 4	30.59	29	68	60.5
L - 5	30.56	28	69	64.0
G - 7	30.48	25	70	73.5
P - 7	30.00	30	71	56.0
N - 3	29.76	24	72	78.0
T - 8	29.64	28	73	64.0
H - 3	29.44	27	74	67.5
G - 5	29.00	29	75	60.5
N - 7	28.74	24	76	78.0
N - 2	28.00	25	77	73.5
G - 1	27.99	27	78	67.5
N - 8	27.94	21	79	81.0
P - 3	27.00	33	80	41.5
L - 7	26.75	25	81	73.5
P - 8	25.88	26	82	70.0
L - 4	23.76	25	83	73.5
G - 2	22.75	22	84	80.0

Appendix D: Pretest 69-A

	W		I		P		U		M		C	
1. Meaningful Learning	25	31%	23	28%	10	12%	3	4%	12	15%	8	10%
2. Transfer of Learning	31	38%	23	28%	11	14%	4	5%	9	11%	3	4%
3. Educational Theories	1	1%	11	14%	7	9%	45	56%	14	17%	3	4%
4. Counseling	21	26%	24	30%	13	16%	0	0%	18	22%	5	6%
5. Creative Thinking	10	12%	12	15%	6	7%	12	15%	29	36%	12	15%
6. Effective Thinking	6	7%	4	5%	10	12%	0	0%	27	33%	35	43%
7. Group Dynamics	21	26%	16	20%	13	16%	2	2%	16	20%	13	16%
8. How We Communicate	31	38%	24	30%	10	12%	4	5%	7	9%	6	7%
9. Plan. for Comm.	41	51%	17	21%	5	6%	0	0%	10	12%	8	10%
10. Plan. for Comm. (Spt)	22	27%	20	25%	9	11%	9	11%	15	19%	9	11%
11. Plan. for Comm. (Spt)	41	51%	17	21%	3	4%	4	5%	12	15%	5	6%
12. Inst. as a Speaker	37	46%	18	22%	12	15%	1	1%	6	7%	8	10%
13. Listening	43	53%	9	11%	9	11%	1	1%	9	11%	10	12%
14. Reading	55	68%	10	12%	6	7%	0	0%	5	6%	5	6%
15. Lesson Planning	15	19%	17	21%	10	12%	7	9%	22	27%	10	12%
16. Selection of Methods	21	26%	17	21%	9	11%	15	18.5%	13	16%	6	7%
17. Dem-Perf. Method	7	9%	9	11%	8	10%	2	2%	32	42%	23	28%
18. Guided Disc Method	23	28%	19	23%	10	12%	7	9%	15	18.5%	6	7%
19. Lecture Method	46	57%	10	12%	8	10%	0	0%	10	12%	5	6%
20. Intro. to Eval. (Char.)	16	20%	10	12%	7	9%	8	10%	23	28%	16	20%
21. Test Const.	23	28%	16	20%	11	14%	7	9%	16	20%	8	10%
22. Test Const.	9	11%	15	18.5%	8	10%	10	12%	22	27%	14	17%
23. Meth. of Grading (T-Score)	16	20%	3	4%	3	2%	50	62%	7	9%	3	4%
24. Item Analysis (D.I.)	3	4%	1	1%	1	1%	61	75%	12	15%	5	6%
25. Performance Rating (errors)	11	14%	13	16%	19	23%	22	27%	9	11%	4	5%

Appendix E: Post Test 69-A

	W		I		P		U		M		C	
1. Man & Relation to Learning	44	64%	3	4%	4	6%	0	0%	4	6%	14	20%
2. Man & Relation to Learning	44	64%	1	1.2%	1	1.2%	1	1.2%	6	9%	16	23%
3. Intro. to Learning	48	70%	6	9%	5	7%	0	0%	3	4%	7	10%
4. Transfer of Learning*	51	74%	5	7%	4	6%	1	1.3%	3	4%	5	7%
5. Transfer of Learning	48	70%	7	10%	3	4%	2	3%	2	3%	7	10%
6. Meaningful Learning*	41	59%	8	12%	6	9%	0	0%	2	3%	12	17%
7. Meaningful Learning	38	55%	6	9%	9	13%	5	7%	3	4%	8	12%
8. Human Relations	51	74%	3	4%	4	6%	0	0%	2	3%	9	13%
9. Group Dynamics*	33	48%	6	9%	8	12%	0	0%	6	9%	16	23%
10. Counseling*	58	84%	2	3%	1	1.3%	0	0%	2	3%	6	9%
11. Counseling	48	70%	6	9%	5	7%	2	3%	4	6%	4	6%
12. Creative Thinking*	26	38%	6	9%	7	10%	4	6%	5	7%	21	30%
13. Creative Thinking	52	75%	2	3%	2	3%	0	0%	3	4%	10	15%
14. Critiquing	38	55%	5	7%	7	10%	2	3%	6	9%	11	16%
15. Critiquing	37	54%	10	15%	8	12%	0	0%	2	3%	12	17%
16. Effective Thinking*	55	80%	1	1.3%	5	7%	0	0%	2	3%	6	9%
17. Effective Thinking	28	41%	1	1.3%	6	9%	1	1.3%	5	7%	28	41%
18. Use of Theories	36	52%	4	6%	3	4%	7	10%	2	3%	17	25%
19. Use of Theories (Essential)*	60	87%	2	3%	0	0%	0	0%	4	6%	3	4%
20. Use of Theories (Perren.)	47	68%	5	7%	7	10%	2	3%	3	4%	5	7%
21. Reading*	58	84%	3	4%	3	4%	0	0%	1	1.3%	4	6%
22. Listening	56	81%	4	6%	6	9%	0	0%	2	3%	1	1.3%
23. Listening*	49	71%	2	3%	3	4%	1	1.3%	3	4%	11	16%
24. Listening	28	41%	10	15%	4	6%	3	4%	6	9%	18	26%
25. Inst. as a Speaker	36	52%	3	4%	4	6%	1	1.3%	1	1.3%	24	35%

Appendix E: Post Test 69-A

	W		I		P		U		M		C	
26. Inst. as a Speaker*	46	67%	6	9%	6	9%	0	0%	0	0%	11	16%
27. Speech to Motivate	48	70%	4	6%	5	7%	0	0%	3	4%	9	13%
28. How We Communicate*	57	83%	5	7%	1	1.3%	0	0%	1	1.3%	5	7%
29. Plan. for Comm. (Organ.)	35	51%	9	13%	12	17%	0	0%	5	7%	8	12%
30. Plan. for Comm. (Organ.)	58	84%	5	7%	1	1.3%	0	0%	1	1.3%	4	6%
31. Plan. for Comm. (patterns)*	57	83%	2	3%	1	1.3%	0	0%	3	4%	6	9%
32. Plan. for Comm. (patterns)	51	74%	5	7%	8	12%	0	0%	1	1.3%	4	6%
33. Plan. for Comm. (Support)	55	80%	5	7%	5	7%	1	1.3%	1	1.3%	2	3%
34. Plan. for Comm. (Support)	42	61%	8	12%	9	13%	0	0%	2	3%	9	12%
35. Plan. for Comm. (Support)*	63	91%	2	3%	0	0%	1	1.3%	0	0%	3	4%
36. Plan. for Comm. (Support)	46	67%	8	12%	8	12%	2	3%	1	1.3%	4	6%
37. Use of TV	47	68%	3	4%	9	13%	1	1.3%	4	6%	5	7%
38. Inst. Aids	58	84%	5	7%	1	1.3%	0	0%	1	1.3%	4	6%
39. Inst. Aids	55	80%	7	10%	3	4%	0	0%	0	0%	4	6%
40. Lesson Planning*	44	64%	6	9%	5	7%	1	1.3%	3	4%	10	15%
41. Lesson Planning	51	74%	8	12%	4	6%	0	0%	1	1.3%	5	7%
42. Selection of Method*	42	61%	4	6%	2	3%	5	7%	1	1.3%	15	22%
43. Teaching Interview	45	65%	5	7%	6	9%	2	3%	4	6%	7	10%
44. Dem-Perf. Method	37	54%	11	16%	7	10%	2	3%	5	7%	7	10%
45. Guided Discussion Method*	49	71%	5	7%	3	4%	0	0%	2	3%	10	15%
46. Guided Discussion Method	17	25%	1	1.3%	8	12%	4	6%	6	9%	33	38%
47. Lecture Method*	61	88%	1	1.3%	4	6%	0	0%	0	0%	3	4%
48. Lecture Method	46	67%	6	9%	7	10%	1	1.3%	3	4%	6	9%
49. Use of the Case	36	52%	5	7%	12	17%	0	0%	6	9%	10	15%
50. Use of Question	47	68%	4	6%	4	6%	2	3%	4	6%	8	12%

Appendix E: Post Test 69-A

	W		I		P		U		M		C	
51. Use of Question	58	84%	5	7%	3	4%	1	1.3%	0	0%	2	3%
52. Program Instruction	27	39%	6	9%	1	1.3%	16	23%	3	4%	16	23%
53. Intro. to Evaluation	41	59%	7	10%	5	7%	2	3%	1	1.3%	13	19%
54. Intro. to Eval. (Char.)*	60	87%	2	3%	1	1.3%	0	0%	1	1.3%	5	7%
55. Intro. to Eval. (Char.)	44	64%	3	4%	5	7%	1	1.3%	1	1.3%	15	22%
56. Intro. to Eval. (Char.)	55	80%	3	4%	2	3%	2	3%	1	1.3%	6	9%
57. Test Const. (Char.)	36	52%	7	10%	10	15%	2	3%	3	4%	11	16%
58. Test Const. (Char.)	32	46%	3	4%	7	10%	3	4%	9	13%	15	22%
59. Test Const.*	59	86%	1	1.3%	4	6%	0	0%	1	1.3%	4	6%
60. Test Const.	48	70%	4	6%	2	3%	0	0%	5	7%	10	15%
61. Test Const.	40	58%	1	1.3%	3	4%	3	4%	9	13%	13	19%
62. Test Const.*	44	64%	6	9%	3	4%	1	1.3%	4	6%	11	16%
63. Method of Grading	57	83%	1	1.3%	0	0%	3	4%	2	3%	6	9%
64. Methods of Grading (R.O.)	37	54%	1	1.3%	1	1.3%	1	1.3%	0	0%	29	42%
65. Methods of Grading (CR)	61	88%	2	3%	1	1.3%	1	1.3%	1	1.3%	3	4%
66. Methods of Grading (S.S.)	62	90%	1	1.3%	0	0%	3	4%	1	1.3%	2	3%
67. Methods of Grading (T-S)	58	84%	1	1.3%	1	1.3%	4	6%	1	1.3%	4	6%
68. Methods of Grading (Raw)*	58	84%	0	0%	1	1.3%	3	4%	2	3%	5	7%
69. Performance Rating (Errors)	31	45%	5	7%	6	9%	3	4%	6	9%	18	26%
70. Performance Rating	58	84%	4	6%	0	0%	1	1.3%	2	3%	4	6%
71. Performance Rating	29	42%	4	6%	6	9%	3	4%	6	7%	22	32%
72. Performance Rating*	42	61%	4	6%	4	6%	4	6%	4	6%	11	16%
73. Test Analysis (EI)	62	90%	1	1.3%	0	0%	0	0%	1	1.3%	5	7%
74. Test Analysis (DI)*	63	92%	0	0%	0	0%	3	4%	0	0%	3	4%
75. Test Analysis	47	68%	2	3%	4	6%	2	3%	5	7%	9	13%
76. Test Analysis	34	49%	3	4%	4	6%	2	3%	3	4%	23	33%

Appendix F: Pretest 69-B

	W		I		P		U		M		C	
1. Meaningful Learning	30	36%	12	14%	20	24%	3	6%	8	9%	11	13%
2. Transfer	38	45%	17	20%	13	15%	7	8%	4	5%	5	6%
3. Theories	5	6%	5	5%	10	12%	42	50%	12	14%	11	13%
4. Counseling	27	32%	9	11%	14	17%	1	1%	16	19%	17	20%
5. Creative Thinking	6	6%	2	2%	8	9%	9	10%	37	44%	23	27%
6. Effective Thinking	2	2%	2	2%	8	9%	2	2%	28	33%	42	50%
7. Group Dynamics	22	26%	13	15%	15	18%	4	5%	13	15%	17	20%
8. How We Communicate	23	27%	13	15%	16	19%	14	17%	13	15%	5	6%
9. Organ. for Comm.	31	37%	12	14%	7	8%	8	9%	12	14%	14	17%
10. Support	26	31%	7	8%	6	7%	11	13%	16	19%	18	21%
11. Support	54	64%	10	12%	5	6%	0	0%	7	8%	8	9%
12. Inst. as a Speaker	31	37%	11	13%	16	19%	2	2%	10	12%	14	17%
13. Speech to Motivate	31	37%	9	11%	16	19%	10	12%	10	12%	8	9%
14. Organ. for Comm.	6	7%	5	6%	25	30%	9	11%	19	22%	20	24%
15. Lesson Planning	16	19%	5	6%	14	17%	8	9%	17	20%	24	29%
16. Selection of Method	19	23%	2	2%	11	13%	20	24%	18	21%	14	17%
17. Dem-Perf.	6	7%	3	4%	7	8%	0	0%	22	26%	46	55%
18. Guided Discussion	35	42%	12	14%	11	13%	3	4%	15	18%	8	9%
19. Lecture	36	43%	7	8%	13	15%	5	6%	11	13%	12	14%
20. Intro. to Evaluation	15	18%	4	5%	5	6%	10	12%	21	25%	29	35%
21. Test Construction	25	30%	10	12%	16	19%	6	7%	16	19%	11	13%
22. Test Construction	4	5%	6	7%	16	19%	6	7%	23	27%	28	33%
23. Method of Grading	7	8%	2	2%	5	6%	58	69%	4	5%	8	9%
24. Analysis (DI)	3	4%	3	4%	1	1%	60	71%	6	7%	11	13%
25. Performance Rating	10	12%	7	8%	13	15%	32	38%	11	13%	11	13%

Appendix G: Post Test 69-B

	W		I		P		U		M		C	
1. Man & Relation to Learning	52	62%	4	5%	5	6%	1	1%	5	6%	17	20%
2. Introduction to Learning	45	54%	3	4%	7	8%	1	1%	4	5%	24	29%
3. Transfer of Learning	68	81%	6	7%	4	5%	0	0%	2	2%	4	5%
4. Basic Concept of Learning	47	56%	7	8%	6	7%	4	5%	7	8%	13	15%
5. Human Relations	50	60%	4	5%	17	20%	0	0%	1	1%	12	14%
6. Counseling	60	71%	5	6%	2	2%	0	0%	3	4%	14	17%
7. Counseling	56	67%	8	10%	9	11%	4	5%	1	1%	6	7%
8. Creative Thinking	17	20%	4	5%	8	10%	3	4%	18	21%	34	40%
9. Critiquing	35	42%	5	6%	11	13%	1	1%	11	13%	21	25%
10. Effective Thinking	28	33%	2	2%	8	10%	1	1%	11	13%	34	40%
11. Educational Theories	29	35%	3	4%	3	4%	7	8%	7	8%	35	42%
12. Educational Theories (Perren.)	37	44%	3	4%	9	11%	7	8%	8	10%	20	24%
13. Reading	66	79%	2	2%	4	5%	1	1%	2	2%	9	11%
14. Listening	25	30%	2	2%	6	7%	1	1%	6	7%	44	52%
15. Listening	9	11%	5	6%	6	7%	2	2%	13	15%	49	58%
16. Inst. as a Speaker	48	57%	1	1%	11	13%	2	2%	6	7%	16	19%
17. Speech to Persuade	58	69%	5	6%	6	7%	2	2%	6	7%	7	8%
18. How We Communicate	27	32%	5	6%	12	14%	2	2%	8	10%	30	36%
19. Organ. for Comm.	62	74%	6	7%	7	8%	0	0%	6	7%	3	4%
20. Organ. for Comm.	63	75%	5	6%	7	8%	0	0%	2	2%	7	8%
21. Use of Support	55	65%	2	2%	7	8%	1	1%	7	8%	12	14%
22. Use of Support	65	77%	4	5%	1	1%	1	1%	0	0%	13	15%
23. Use of TV	50	59%	3	4%	7	8%	1	1%	3	4%	20	24%
24. Instructional Aids	75	89%	5	6%	2	2%	0	0%	1	1%	1	1%
25. Lesson Planning	48	57%	8	10%	3	4%	0	0%	3	4%	22	26%

Appendix G: Post Test 69-B

	W		I		P		U		M		C	
26. Lesson Planning	55	65%	5	6%	3	4%	3	4%	2	2%	16	19%
27. Selection of Method	30	36%	1	1%	8	10%	3	4%	9	11%	33	39%
28. T.I. Method	49	58%	4	5%	11	13%	0	0%	3	4%	17	20%
29. Dem-Perf. Method	47	56%	7	8%	16	19%	0	0%	0	0%	14	17%
30. Guided Discussion Method	71	85%	5	6%	0	0%	0	0%	2	2%	6	7%
31. Lecture Method	55	65%	5	6%	8	10%	4	5%	3	4%	9	11%
32. Case Study Method	40	48%	8	10%	15	18%	0	0%	6	7%	15	18%
33. Use of Questions	72	86%	3	4%	3	4%	0	0%	2	2%	4	5%
34. Programmed Instruction	18	21%	4	5%	8	10%	36	43%	5	6%	13	15%
35 Introduction to Evaluation	39	35%	2	2%	7	8%	4	5%	11	13%	31	37%
36. Intro. to Eval. (Char.)	27	32%	4	5%	5	6%	2	2%	9	11%	37	44%
37. Intro. to Eval. (Char.)	54	64%	2	2%	3	4%	4	5%	2	2%	19	23%
38. Test Const. (Char.)	41	49%	2	2%	8	10%	2	2%	6	7%	25	30%
39. Test Const.	65	77%	2	2%	7	8%	1	1%	2	2%	7	8%
40. Test Const.	41	49%	5	6%	9	11%	4	5%	4	5%	21	25%
41. Method of Grading	55	65%	1	1%	2	2%	4	5%	2	2%	20	24%
42. Method of Grading (CR)	62	74%	1	1%	4	5%	1	1%	3	4%	13	15%
43. Method of Grading (S.S.)	61	73%	2	2%	5	6%	8	10%	0	0%	8	10%
44. Method of Grading (T)	48	57%	0	0%	3	4%	19	23%	3	4%	11	13%
45. Performance Rating	58	69%	2	2%	3	4%	2	2%	3	4%	16	19%
46. Performance Rating	22	26%	1	1%	8	19%	6	7%	7	8%	40	48%
47. Performance Rating	40	48%	5	6%	11	13%	6	7%	4	5%	17	20%
48. Item Analysis (EI)	56	67%	0	0%	0	0%	10	12%	5	6%	13	15%
49. Item Analysis (DI)	56	67%	0	0%	4	5%	8	10%	6	7%	10	12%
50. Item Analysis (General)	44	52%	4	5%	3	4%	8	10%	5	6%	20	24%

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified

ORIGINATING ACTIVITY (Corporate author) The Shuford-Massengill Corporation One Wallis Court Lexington, Massachusetts 02173	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

REPORT TITLE
The Use of Confidence Testing in The Academic Instructor Course

DESCRIPTIVE NOTES (Type of report and inclusive dates)
Scientific Interim

AUTHOR(S) (First name, middle initial, last name)
Willie C. Gardner, Jr.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
February 1970	56	7

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
F44620-69-C-0068 (ARPA) b. PROJECT NO. 9719 61101D c. 681313	SMC R-20 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFOSR-70-0143-TR

10. DISTRIBUTION STATEMENT
1. This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
TECH, OTHER	Air Force Office of Scientific Research (SRLB) 1400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

The problems stated at the outset were: to discover whether information provided by confidence testing would result in a more accurate assessment of student knowledge so more personal individualized instruction could be presented, and whether sufficient precise data could be attained to be of significant aid in curriculum planning. Analysis has shown that much qualitative and quantitative information is provided when employing the methods and materials associated with confidence testing.

Predictive validity appears to be positive, but not significantly high enough to forecast a student's performance on practice teaching lessons or his final standing in the class. The ***assessment of student knowledge*** tends to be more accurate when using confidence tests. ***Reliability*** of test instruments seems to be improved when using confidence tests. Confidence tests provide more ***information*** about the students than do choice tests. Better ***item analysis data*** is provided with confidence tests than with the standard choice tests.

DD FORM 1473

UNCLASSIFIED
Security Classification

KEY WORDS						
Confidence testing						
Educational measurement						
Item analysis						
Multiple-choice tests						
Reliability						
Validity						